Dynamics of random recurrent networks with correlated low-rank structure

Friedrich Schuessler [1], Alexis Dubreuil,[2] Francesca Mastrogiuseppe,[3] Srdjan Ostojic,[2] and Omri Barak[1,*]

[1]*Rappaport Faculty of Medicine and Network Biology Research Group, Technion–Israel Institute of Technology, Haifa, Israel*
[2]*Laboratoire de Neurosciences Cognitives et Computationnelles, INSERM U960,*
Ecole Normale Superieure - PSL Research University, 75005 Paris, France
[3]*Gatsby Computational Neuroscience Unit, UCL, London, United Kingdom*

A given neural network in the brain is involved in many different tasks. This implies that, when considering a specific task, the network's connectivity contains a component which is related to the task and another component which can be considered random. Understanding the interplay between the structured and random components and their effect on network dynamics and functionality is an important open question. Recent studies addressed the coexistence of random and structured connectivity but considered the two parts to be uncorrelated. This constraint limits the dynamics and leaves the random connectivity nonfunctional. Algorithms that train networks to perform specific tasks typically generate correlations between structure and random connectivity. Here we study nonlinear networks with correlated structured and random components, assuming the structure to have a low rank. We develop an analytic framework to establish the precise effect of the correlations on the eigenvalue spectrum of the joint connectivity. We find that the spectrum consists of a bulk and multiple outliers, whose location is predicted by our theory. Using mean-field theory, we show that these outliers directly determine both the fixed points of the system and their stability. Taken together, our analysis elucidates how correlations allow structured and random connectivity to synergistically extend the range of computations available to networks.

I. INTRODUCTION

One of the central paradigms of neuroscience is that computational function determines connectivity structure: if a neural network is involved in a given task, its connectivity must be related to this task. However, a given circuit's connectivity also depends on development and the learning of a multitude of tasks [1,2]. Accordingly, connectivity has often been depicted as containing a sum of random and structured components [3–7]. Given that structure emerges through adaptive processes on top of existing random connectivity, one would intuitively expect correlations between the two components. Nevertheless, the functional effects of the interplay between the random and the structured components have not been fully elucidated.

Networks designed to solve specific tasks often use purely structured connectivity [8–10] that has been analytically dissected [11]. The dynamics of networks with purely random connectivity were also thoroughly explored, charting the transitions between chaotic and ordered activity regimes [12–17]. Adding *uncorrelated* random connectivity to a structured one was shown to generate the activity statistics originating from the random component while retaining the functional aspects of the structured one [4–6,18].

A specific setting in which correlations between random and structured components arise is the training of initially random networks to perform tasks. One class of training algorithms, reservoir computing, only modifies a feedback loop on top of the initial random connectivity [19–21]. These algorithms can be used to obtain a wide range of computations [22,23]. Recently, a specific instance of a network trained to exhibit multiple fixed points was analytically examined [3]. It was shown that the dependence between the feedback loop and the initial connectivity is essential to obtain the desired functionality, but the explicit form of the correlations and the manner in which they determine functionality remained elusive.

Thus there is no general theory linking the correlations between random and structured components to network dynamics. Here we address this issue by examining the nonlinear dynamics of networks with such correlations. Because the dynamics of nonlinear systems vary between different areas of phase space, we focus on linearized dynamics around different fixed points. To facilitate the analysis, we consider low-rank structured components which were shown to allow for a wide range of functionalities [4].

We develop a mean-field theory that takes into account correlations between the random connectivity and the low-rank part. Our theory directly links these correlations to the spectrum of the connectivity matrix. We show how a correlated rank-one perturbation can lead to multiple spectral outliers and fixed points, a phenomenon that requires high-rank perturbations in the uncorrelated case [4]. We analytically

study dynamics around nontrivial fixed points, revealing a surprising connection between the spectrum, the fixed points and their stability. Taken together, we show how correlations between the low-rank structure and the random connectivity extend the computations of the joint network beyond the sum of its parts.

II. NETWORK MODEL

We examine the dynamics of recurrent neural networks with correlated random and structured components in their connectivity. The structured component P is a low-rank matrix and the random component J is a full-rank matrix. Network dynamics with such a connectivity structure have been analyzed for P being independent of the random connectivity [4]. The learning frameworks of echo state networks and FORCE also have such connectivity structure [20,21]. There, however, the structure P is trained such that the full network performs a desired computation, possibly correlating P to J.

For most of this study, we set the rank of P to one and write it as the outer product

$$P = \mathbf{m}\mathbf{n}^T \tag{1}$$

of the two structure vectors $\mathbf{m}$ and $\mathbf{n}$. The matrix J and vector $\mathbf{m}$ are drawn independently from normal distributions, $J_{ij} \sim \mathcal{N}(0, g^2/N)$ and $m_i \sim \mathcal{N}(0, 1)$, where N is the network size and g controls the strength of the random part [12]. The second vector $\mathbf{n}$ is defined in terms of J and $\mathbf{m}$. In this sense, $\mathbf{n}$ carries the correlation between J and P. This is in line with the echo state and FORCE models, where $\mathbf{n}$ corresponds to the readout vector which is trained and therefore becomes correlated to J and $\mathbf{m}$. In contrast to these models, however, we constrain the statistics of $\mathbf{n}$ to be Gaussian. This allows for an analytical treatment and thus for a transparent understanding of how the correlations affect the network dynamics.

The details of the construction of $\mathbf{n}$ are described later on. At this point, we merely state that the entries of $\mathbf{n}$ scale with the network size as $1/N$. The structure P is hence considered as a perturbation to the random connectivity J whose entries scale as $1/\sqrt{N}$. All our results are valid in the limit of infinitely large networks, $N \to \infty$. Throughout the work, we compare the theoretical predictions with samples from finite networks.

The network dynamics are given by standard rate equations. Neurons are characterized by their internal states x_i and interact with each other via firing rates $\phi(x_i)$. The nonlinear transformation from state to firing rate is taken to be the hyperbolic tangent, $\phi = \tanh$. The entire network dynamics are written as

$$\dot{\mathbf{x}}(t) = -\mathbf{x}(t) + (J + P)\phi(\mathbf{x}(t)), \tag{2}$$

with the state vector $\mathbf{x} \in \mathbb{R}^N$ and the nonlinearity applied elementwise. The derivation of our results, Appendix E, further includes a constant external input $\mathbf{I}$. The results in the main text, however, only consider the autonomous network.

III. LINEAR DYNAMICS AROUND THE ORIGIN

The origin $\mathbf{x} = \mathbf{0}$ is a fixed point, since $\phi(0) = 0$. It is stable if the real parts of all the eigenvalues of the Jacobian

are smaller than one. Since $\phi'(0) = 1$, the Jacobian is simply the connectivity matrix $J + \mathbf{m}\mathbf{n}^T$ itself. Here we examine the spectral properties of this matrix.

A. Eigenvalues

The spectrum of the Gaussian random matrix J converges to a uniform distribution on a disk with radius g and centered at the origin for $N \to \infty$ [24]. Previous studies have explored the effect of independent low-rank perturbations like in our model [25,26]. They found that the limiting distribution of the remaining eigenvalues, referred to as the bulk, does not change. Additionally, the spectrum contains outliers corresponding to the eigenvalues of the low-rank perturbation itself. In this sense, the spectra of the random matrix J and the low-rank perturbation decouple (although the precise location of each eigenvalue is affected by the perturbation). To our knowledge, the effect of correlated low-rank perturbations, which we explore below, has not been considered before.

To determine the spectrum, we apply the matrix determinant lemma [27]:

$$\det(A + \mathbf{m}\mathbf{n}^T) = (1 + \mathbf{n}^T A^{-1}\mathbf{m})\det(A), \tag{3}$$

where $A \in \mathbb{C}^{N \times N}$ is an invertible matrix. For a complex number z that is not an eigenvalue of J, the matrix $J - \mathbb{1}z$ is invertible, resulting in

$$\det((J + \mathbf{m}\mathbf{n}^T) - \mathbb{1}z)$$
$$= (1 + \mathbf{n}^T(J - \mathbb{1}z)^{-1}\mathbf{m})\det(J - \mathbb{1}z). \tag{4}$$

The roots of this equation are the eigenvalues of $J + \mathbf{m}\mathbf{n}^T$. Since the determinant on the right-hand side is nonzero, we get the scalar equation

$$z = \mathbf{n}^T\left(\mathbb{1} - \frac{J}{z}\right)^{-1}\mathbf{m}. \tag{5}$$

As long as the entire spectrum is affected by the rank 1 perturbation, this equation determines all eigenvalues of $J + \mathbf{m}\mathbf{n}^T$. We are interested in outliers of the spectrum: eigenvalues of $J + \mathbf{m}\mathbf{n}^T$ larger than the spectral radius of J (which in the limit of $N \to \infty$ is given by g). For such an outlier, denoted by λ, the inverse in Eq. (5) can be written as a series, and we have

$$\lambda = \sum_{k=0}^{\infty} \frac{\theta_k}{\lambda^k}, \tag{6}$$

with the overlaps

$$\theta_k = \mathbf{n}^T J^k \mathbf{m}. \tag{7}$$

Although this equation is a polynomial of infinite degree, there can be at most proper N solutions (those outside of the bulk; see Appendix A).

The series representation Eq. (6) is the main result of this section. It indicates that the overlaps θ_k between $\mathbf{m}$ and $\mathbf{n}$ after passing through J for k times determine the eigenvalues of the perturbed matrix. It is hence useful to characterize the correlations between J and the rank-one perturbation in terms of these overlaps.

The description up to this point is general and does not depend on details of the matrix J. For our model, where J is a
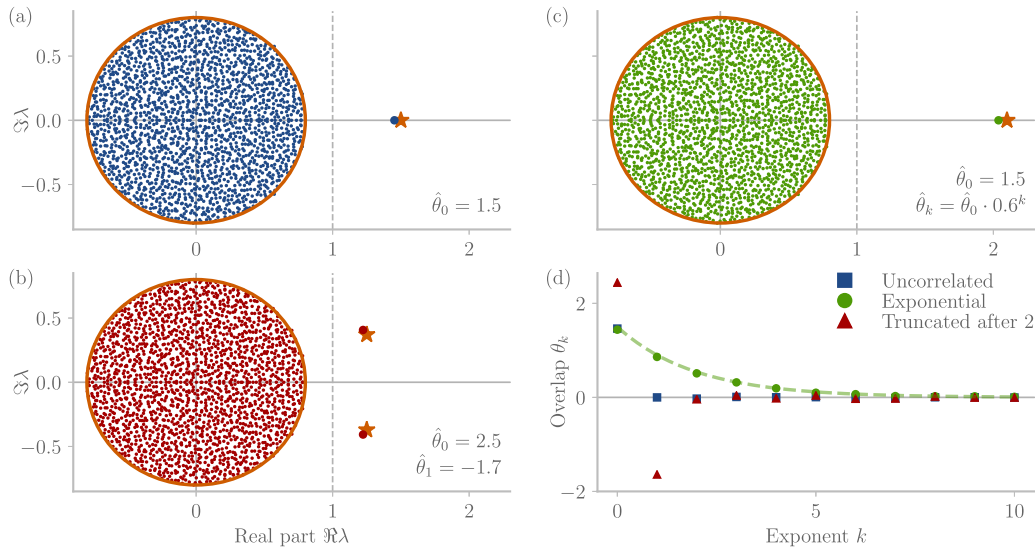
FIG. 1. Spectral outliers via low-rank perturbations. Spectrum of $J + \mathbf{m}\mathbf{n}^T$ with (a) no correlations, (b) exponential overlaps, and (c) truncated overlaps, Eq. (11). See Eq. (B1) for details on the construction of $\mathbf{n}$. The values of nonzero $\hat{\theta}_k$ are displayed in each plot. Orange circles and stars indicate the theoretical prediction, dots refer to the spectra of the finite-size connectivity matrices, computed numerically. (d) Overlaps $\theta_k = \mathbf{n}^T J^k \mathbf{m}$ for the cases above. The dashed line are the target overlaps $\hat{\theta}_k$ for the exponential correlation. Parameters: $N = 2000$ and $g = 0.8$.

random matrix, the scalar products θ_k over N entries are self-averaging: for $N \to \infty$, θ_k converges to its ensemble average $\mathbb{E}[\theta_k]$, with variance decaying as $1/N$. We rely on this property and compute quantities for single realization of large networks instead of ensemble averages.

A random matrix J has the effect of decorrelating independent vectors: if the vectors $\mathbf{m}$ and $\mathbf{n}$ are uncorrelated to J, a single pass through the network already annihilates any overlap between $\mathbf{n}$ and $J\mathbf{m}$. In Appendix B, we formally show that the self-averaging indeed yields $\mathbf{n}^T J^k \mathbf{m} = 0$ for $k \geqslant 1$. We can apply this to Eq. (6): if $\theta_k = 0$ for any $k \geqslant 1$, then

$$\lambda = \theta_0 = \mathbf{n}^T \mathbf{m}. \tag{8}$$

Thus an independent rank-one perturbation yields a single outlier positioned at the eigenvalue of the rank-one matrix itself [Fig. 1(a)], in accordance with known results [25,26].

If $\mathbf{m}\mathbf{n}^T$ is correlated to J, the θ_k will not vanish for nonzero k. We analyze two special cases.

(i) If $\theta_k = 0$ for any $k \geqslant 2$, then there are two outliers

$$\lambda_\pm = \frac{\theta_0}{2} \pm \sqrt{\left(\frac{\theta_0}{2}\right)^2 + \theta_1}. \tag{9}$$

This can give rise to complex conjugate outliers, as displayed in Fig. 1(b). More generally, K nonzero overlaps lead to K outliers via a polynomial equation [Eq. (C1)].

(ii) A second case is one of a converging series in Eq. (6). The simplest assumption is an exponential scaling, $\theta_k = \theta_0 b^k$ with base b. Inserting into the eigenvalue equation (6) yields a single solution

$$\lambda = \theta_0 + b. \tag{10}$$

Remarkably, we see that correlation between the random matrix J and the rank-one perturbation does not necessarily

lead to more than one outlier. This is shown in Fig. 1(c). The observation generalizes to correlations expressible as a sum of K exponentially decaying terms, leading to K outliers [Eq. (C3)].

We can apply this understanding to construct a network with a set of outliers and either one of the underlying correlation structures. One way is to define the vector $\mathbf{n}$ explicitly in terms of $\mathbf{m}$ and J. For example, if we set

$$\mathbf{n} = \frac{1}{N}\left(\hat{\theta}_0 \mathbf{m} + \frac{\hat{\theta}_1}{g^2} J\mathbf{m}\right), \tag{11}$$

then the overlaps will self-average to $\mathbb{E}[\theta_k] = \hat{\theta}_k$ for $k \in \{0, 1\}$ and $\mathbb{E}[\theta_k] = 0$ for any $k \geqslant 2$, with variance scaling as $1/N$. This is shown formally and generalized to higher θ_k in Appendix B. The details of the construction for a set of target outliers is further detailed in Appendix C. The discrepancy between numerical and target outliers in Fig. 1 is due to finite size effects, which decay with $1/\sqrt{N}$ (verified numerically, and in accordance with Ref. [4]).

The simulations further show that the remaining eigenvalues span the same circle as without the perturbation. While all eigenvalues change, visual inspection does not reveal any changes in the statistics.

B. Implementation of multiple outliers

So far we analyzed the outliers for given correlations between J and $\mathbf{m}\mathbf{n}^T$ as quantified by the overlaps θ_k. We now change the perspective and ask about the properties of the rank-one perturbation given a set of outliers. We saw that in principle a given set of outliers may have multiple underlying correlation structures, e.g., through a truncated set of nonzero overlaps or a combination of exponentially decaying terms. Regardless of the correlation structure, however, we observe that the norm of $\mathbf{n}$ grows fast with the number of outliers
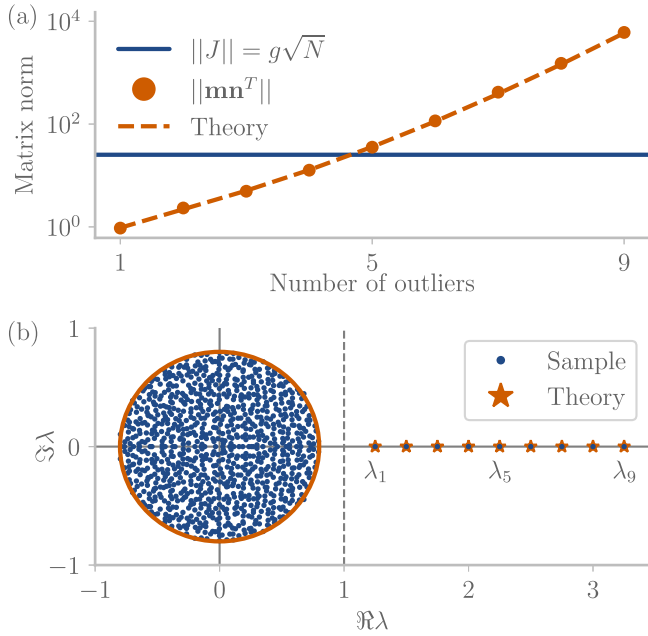
FIG. 2. Scaling of the norm of the rank-one perturbation with number of induced outliers. The vector $\mathbf{n}$ is the least square solution to implementing a set of outliers $\Lambda = \{\lambda_1, \ldots, \lambda_K\}$, with $\lambda_k = 1.25 + 0.25k$, see Appendix D. (a) Log-linear plot of the Frobenius norm of J and $\mathbf{mn}^T$ as a function of the number of outliers. The dashed line is the theoretical prediction. (b) Spectrum of $J + \mathbf{mn}^T$ for $K = 9$ outliers. Parameters: $N = 1000$ and $g = 0.8$.

introduced, implying that strong perturbations are needed to generate a large number of outliers.

To understand analytically the origin of this phenomenon, we focus on a method to determine the least square $\mathbf{n}$ given J, $\mathbf{m}$, and the set of target outliers Λ. The resulting $\mathbf{n}$ can be formulated using the pseudoinverse, as detailed in Appendix D. The main result of this analysis is the scaling of the Frobenius norm of the rank-one matrix $\mathbf{mn}^T$ with the number of outliers. The asymptotic behavior is given by

$$||\mathbf{mn}^T|| \sim g \prod_{\lambda \in \Lambda} \frac{|\lambda|}{g}, \qquad (12)$$

that is, exponentially growing with the number of outliers. In comparison, the Frobenius norm of J is given by $||J|| = g\sqrt{N}$. This means that if one aims to place more than a handful of outliers, the perturbation $\mathbf{mn}^T$ becomes the dominating term (for a fixed network size N). We illustrate this in Fig. 2 by plotting $||\mathbf{mn}^T||$ for sets of outliers $\Lambda_K = \{\lambda_1, \ldots, \lambda_K\}$ with growing number K. The outliers λ_k were placed on the real line. Further tests including complex eigenvalues gave similar results (not shown). A similar method of deriving $\mathbf{n}$ from the pseudoinverse has been described in Ref. [28].

The scaling (12) shows another important point: the bulk radius g critically determines the norm of the rank-one perturbation. Indeed, the contribution of each outlier λ_k is relative to the radius. Even for a single outlier, where

$$||\mathbf{mn}^T|| = \sqrt{\lambda^2 - g^2}, \qquad (13)$$

an increase in g leads to decreasing norm. This observation suggests that a large random connectivity facilitates the control of the spectrum by a rank-one perturbation.

IV. NONTRIVIAL FIXED POINTS

We now turn to the nontrivial fixed points of the network. At these, the internal states $\mathbf{x}$ obey the equation

$$\mathbf{x} = J\boldsymbol{\phi} + \kappa\mathbf{m}. \qquad (14)$$

Here we defined the scalar feedback strength $\kappa = \mathbf{n}^T\boldsymbol{\phi}$, using the vector notation $\boldsymbol{\phi} = \phi(\mathbf{x})$.

The fixed points of related models have been analyzed in previous works. For infinitely large networks, the unperturbed system ($P = 0$) has a single fixed point at the origin if $g < 1$ [15]. For $g > 1$, the system exhibits chaotic dynamics [12]. In this regime, the number of (unstable) fixed points scales exponentially with the network size N [15]. Here we only focus on networks in the nonchaotic regime, where either $g < 1$ or the perturbation P suppresses chaos [4].

A. Fixed point manifold

Following Rivkind and Barak [3], the perturbed system with fixed points (14) can be understood using a surrogate system in which the feedback κ is replaced with a fixed scalar $\hat{\kappa}$. For $g < 1$, every such value $\hat{\kappa}$ corresponds to a unique fixed point

$$\hat{\mathbf{x}} = J\phi(\hat{\mathbf{x}}) + \hat{\kappa}\mathbf{m}. \qquad (15)$$

This equation defines the one-dimensional nonlinear manifold

$$\mathcal{M} = \{\hat{\mathbf{x}} \,|\, \hat{\kappa} \in \mathbb{R}\}. \qquad (16)$$

The manifold $\mathcal{M}$ can be understood by looking at the asymptotics. For large input $\hat{\kappa}$, the nonlinearity saturates and the manifold becomes approximately linear:

$$\hat{\mathbf{x}}_\sim = \mathbf{c} + \hat{\kappa}\mathbf{m}, \qquad (17)$$

with $\mathbf{c} = J \text{sign}(\mathbf{m})$. Around the origin, we linearize and obtain

$$\hat{\mathbf{x}} = \hat{\kappa}\,\mathbf{a} + \mathcal{O}(\hat{\kappa}^2), \qquad (18)$$

with $\mathbf{a} = (\mathbb{1} - J)^{-1}\mathbf{m}$.

Applying orthonormalization to the triplet $(\mathbf{m}, \mathbf{c}, \mathbf{a})$, we obtain a three-dimensional basis. We observe that, for $N \to \infty$, the vectors $\mathbf{m}$ and $\mathbf{c}$ are orthogonal and that the vector $\mathbf{a}$ becomes orthogonal to the other two in the limit $g \to 1$. Accordingly, we name the coefficients of the basis $(y_\mathbf{m}, y_\mathbf{c}, y_\mathbf{a})$. The projection of the manifold $\mathcal{M}$ on this basis is shown in Fig. 3(c) for three different values of g. Numerical evaluation of the reconstruction error shows that these three dimensions reconstruct the manifold very well albeit with decreasing accuracy for increasing g (not shown).

Fixed points of the full system are obtained by determining κ self-consistently. They necessarily lie on the manifold $\mathcal{M}$. One consequence is a strong correlation between pairs of fixed points, especially if both lie close to the origin or in the saturating regime. In Figs. 3(b)–3(d), we numerically evaluate this correlation for three different randomness strengths g. On
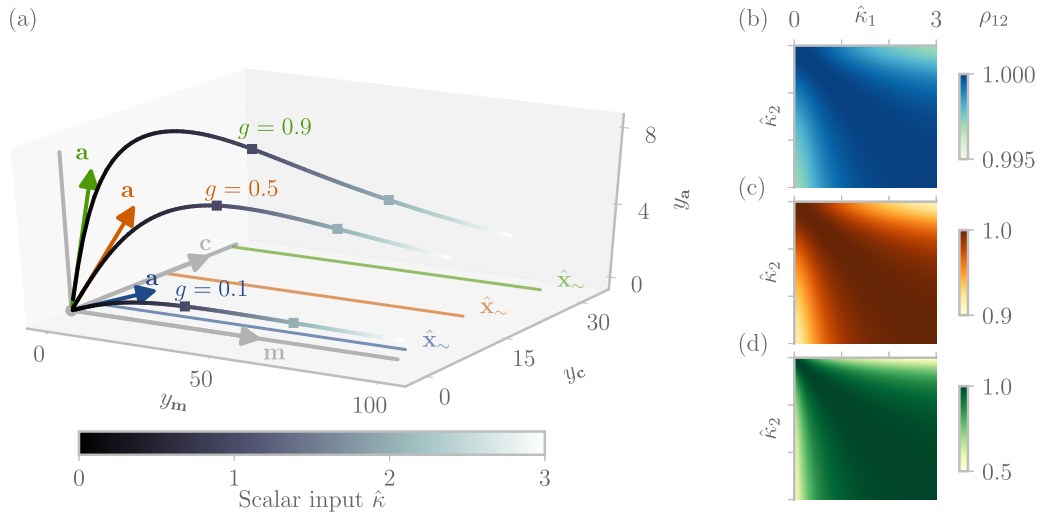
FIG. 3. Manifold $\mathcal{M}$, Eq. (16) constraining fixed points, Eq. (15). (a) Projection of $\mathcal{M}$ for three networks with different strength of randomness g (see main text for the three-dimensional basis). The negative side for $\hat{\kappa} < 0$ is symmetric and not shown. The squares on the manifolds indicate the inputs $\hat{\kappa} = (1, 2)$. The straight lines in the plane $y_{\mathbf{a}} = 0$ are the asymptotic directions $\hat{\mathbf{x}}_\sim$ for the manifolds. [(b)–(d)] Correlation ρ_{12} between two points $\hat{\mathbf{x}}(\hat{\kappa}_i)$ on $\mathcal{M}$ for two different inputs $\hat{\kappa}_1, \hat{\kappa}_2 \in [0, 3]$. (b), (c), and (d) correspond to the random connectivity strengths $g = 0.1, 0.5$, and 0.9, respectively. Note the different scales on the color bars.

can observe that for $g \leqslant 0.5$, correlation does not drop below 90%. Even for $g = 0.9$, the correlation is low only if one fixed point is very close to the origin and the other one is far out.

So far we only considered the case $g < 1$. For $g > 1$, there is a minimal $\hat{\kappa}_{\min}$ for which the dynamics are stabilized and a unique stable fixed point emerges [4]. Here, the manifold $\mathcal{M}$ is unconnected and now reads $\mathcal{M} = \{\hat{\mathbf{x}} \,|\, \hat{\kappa} \in \mathbb{R} \setminus (-\hat{\kappa}_{\min}, \hat{\kappa}_{\min})\}$.

Finally we note that the constraints of a one-dimensional manifold are general and do not depend on the details of the vector $\mathbf{n}$, especially not on its Gaussian statistics. This is particularly important for learning algorithms like the echo state framework or FORCE, which by construction only allow for the adaptation of the vector $\mathbf{n}$ [20,21]. Accordingly, fixed points in these cases are also strongly correlated, which may lead to catastrophic forgetting when trying to learn multiple fixed points sequentially [29].

B. Mean-field theory

For nontrivial fixed points of the full network, Eq. (14), the scalar feedback κ needs to be consistent with the firing rates $\phi(\mathbf{x})$. Similar to prior works, we compute κ using a mean-field theory [4]. The central idea of the mean-field theory is to replace the input to each variable x_i by a stochastic variable with statistics matching the original system. The statistics of the resulting stochastic processes x_i are then computed self-consistently.

Because our model includes correlations between the random part J and the low-rank structure P, the correlations in the activity do not vanish as dynamics unfold. This phenomenon prevents the application of previous theories [4]. We hence develop a new theory. The details are elaborated in Appendix E. Here we give an outline of the analysis.

The starting point is the scalar feedback κ. The Gaussian statistics of $\mathbf{n}$ and the fixed point $\mathbf{x}$ allow to factor out the

effect of the nonlinearity via partial integration. We have

$$\kappa = \mathbf{n}^T \boldsymbol{\phi} = \langle \phi' \rangle \, \mathbf{n}^T \mathbf{x}, \tag{19}$$

with the average slope $\langle \phi' \rangle$ evaluated at the fixed point:

$$\langle \phi' \rangle = \int \mathcal{D}z \, \phi'(\sqrt{\Delta^0} z), \tag{20}$$

where $\mathcal{D}z$ is the standard Gaussian measure. Δ^0 is the variance of $\mathbf{x}$, which from the fixed point equation (14) is given by

$$\Delta^0 = g^2 \langle \phi^2 \rangle + \kappa^2. \tag{21}$$

The quantities $\langle \phi' \rangle$, Δ^0, and κ are determined self-consistently. To that end, we further evaluate κ in Eq. (19). Inserting the fixed point equation (14) yields

$$\mathbf{n}^T \mathbf{x} = \mathbf{n}^T J \boldsymbol{\phi} + \kappa \mathbf{n}^T \mathbf{m}. \tag{22}$$

The first term on the right-hand side vanished in previous studies with no correlation between P and J [4]. In our case, there are correlations, and we proceed to analyze this term. We first interpret $J^T \mathbf{n}$ as a Gaussian vector and use partial integration to replace $\boldsymbol{\phi}$ with $\mathbf{x}$:

$$\mathbf{n}^T \mathbf{x} = \langle \phi' \rangle \mathbf{n}^T J \mathbf{x} + \kappa \mathbf{n}^T \mathbf{m}. \tag{23}$$

We now insert the fixed point equation (14) into the new first term of the right-hand side. We can apply this scheme recursively and arrive at an equation linear in κ on both sides:

$$\kappa = \kappa \langle \phi' \rangle \sum_{k=0}^{\infty} \langle \phi' \rangle^k \theta_k, \tag{24}$$

with overlaps as defined above, Eq. (7). We are looking at a nontrivial fixed point, so we can divide by the nonzero κ to obtain

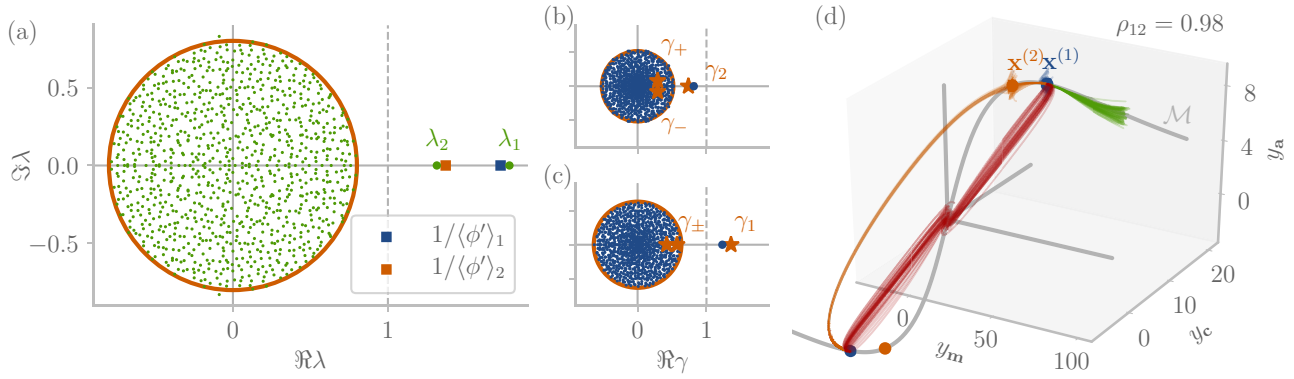$$\frac{1}{\langle \phi' \rangle} = \sum_{k=0}^{\infty} \langle \phi' \rangle^k \theta_k. \tag{25}$$

FIG. 4. Two fixed points induced by rank-one perturbation correlated to the random connectivity J. (a) Spectrum of $J + \mathbf{mn}^T$. The squares indicate the corresponding averaged slopes at the fixed points, as predicted by Eq. (26). [(b) and (c)] Eigenvalues of the Jacobian at the two fixed points $\mathbf{x}^{(1)}$ (b) and $\mathbf{x}^{(2)}$ (c). Stars indicate the theoretical predictions for exceptional stability eigenvalues (only meaningful outside the bulk). (d) Fixed points and manifold $\mathcal{M}$. The colored lines indicate trajectories starting from the two fixed points (blue and orange), a point on the manifold $\mathcal{M}$ (green) and the origin. All trajectories converge on $\mathbf{x}^{(1)}$ or its negative counterpart. At each point, 50 different initial conditions are obtained by adding Gaussian noise (SD = 0.5). The fixed point correlation is indicated by ρ_{12}. Parameters: $g = 0.8$ and $N = 1000$. The rank-one perturbation is obtained by the least-squared $\mathbf{n}$, see Appendix D.

A comparison with Eq. (6) shows that the two equations are identical if

$$\lambda = 1/\langle \phi' \rangle. \tag{26}$$

This is a remarkable relationship between the outliers and autonomously generated fixed points: each nontrivial fixed point $\mathbf{x}^{(i)}$ must be associated with a real eigenvalue λ_i such that the average over the derivative of firing rates at this fixed point, $\langle \phi' \rangle_i$, fulfills the above condition (26). In the special case of $\phi = \tanh$, the $\langle \phi' \rangle_i$ are confined to the interval (0, 1], so the corresponding eigenvalues must be real and larger than one. One may hence look at the spectrum of the connectivity matrix alone and determine how many nontrivial fixed points there are. An instance of this phenomenon is illustrated in Fig. 4. The spectrum at the origin contains two outliers λ_i, $i = 1, 2$, each real and larger than one. The dynamics have two corresponding fixed points $\mathbf{x}^{(i)}$ located on the manifold $\mathcal{M}$. In accordance with Eq. (26), the average slopes at these fixed points, $1/\langle \phi' \rangle_i$, agree with the outliers up to deviations due to the finite network size.

C. Stability of fixed points

The stability of each fixed point is determined by the spectrum of its Jacobian. The associated stability matrix (the Jacobian shifted by $-\mathbb{1}$) is

$$S = (J + \mathbf{mn})R', \tag{27}$$

with the diagonal matrix of slopes $R'_{ij} = \delta_{ij}\phi'_i$. Previous work ([4]) has found that the spectrum of S, too, consists of a bulk and a small number of exceptional eigenvalues: in the case of an uncorrelated rank-one perturbation, there are two nonzero eigenvalues obtained via mean-field theory, only one of which has been found outside the random bulk. The radius of the bulk shrinks to $g\sqrt{\langle \phi'^2 \rangle}$ due to the saturation of the nonlinearity [4]. We find numerically that the bulk behaves alike in our model, too. For the rest of this section, however,

we focus on exceptional eigenvalues of the stability matrix S, denoted by γ.

Similar to the trivial fixed point, one can apply the matrix determinant lemma to derive an equation for the stability eigenvalues:

$$\gamma = \mathbf{n}^T R' \left(\mathbb{1} - \frac{JR'}{\gamma} \right)^{-1} \mathbf{m}. \tag{28}$$

We can apply the mean-field theory introduced above to evaluate the right-hand side. The details of this calculation are deferred to Appendix F. It turns out that the resulting γ are surprisingly compact. We now describe these stability eigenvalues.

Consider the fixed point $\mathbf{x}^{(i)}$. According to Eq. (26), this fixed point corresponds to the eigenvalue λ_i. Eq. (28) always has two solutions $\gamma_\pm$ determined by a quadratic equation. These are only dependent on the outlier λ_i and the statistics of the fixed point $\mathbf{x}^{(i)}$, but entirely independent of the remaining spectrum or other fixed points. Their precise values are detailed in Eq. (G2). It turns out that γ_+ and γ_- always have a real part smaller than one. They hence do not destabilize the fixed point. Additionally, at least one of the two is always hidden within the bulk of the eigenvalues, as observed before for the case of no correlation [4]. In Figs. 4(b) and 4(c), the spectra of the Jacobian at two fixed points are compared with the theoretical predictions. In both cases, $\gamma_\pm$ correspond to the two stars within the bulk. In Fig. 5(b), the bulk is smaller ($g = 0.6$) and γ_+ is visible.

If $\lambda_i = \lambda_1$ is the only outlier, then $\gamma_\pm$ are the only two solutions of Eq. (28), and the fixed point $\mathbf{x}^{(1)}$ as well as its mirror $-\mathbf{x}^{(1)}$ will be stable. However, if there are $K \geqslant 2$ outliers $\{\lambda_1, \ldots, \lambda_K\}$, we find an additional set of $K - 1$ stability eigenvalues

$$\gamma_j = \frac{\lambda_j}{\lambda_i} \quad \text{for all} \quad j \in \{1, \ldots, K\}, \ j \neq i. \tag{29}$$

This expression indicates a remarkable relationship between different fixed points: the existence of a fixed point $\mathbf{x}^{(j)}$ with
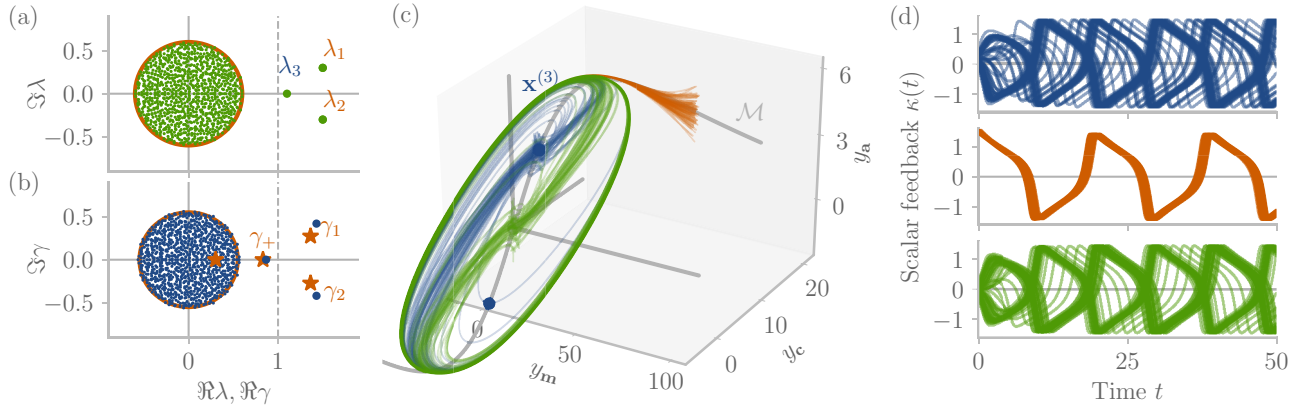
FIG. 5. Limit cycle induced by oscillatory instability. (a) Spectrum of the connectivity matrix. The outlier λ_3 is real-valued and larger than one, so there is a corresponding fixed point $\mathbf{x}_3$. (b) The spectrum at the fixed point. The predicted stability eigenvalue γ_- lies inside the bulk and is not labeled. (c) Fixed point and dynamics as in Fig. 4. Trajectories start at the fixed point (blue), a point on the manifold $\mathcal{M}$ (orange) and the origin (green). (d) Scalar feedback $\kappa(t)$ for the different initial conditions. Parameters: $g = 0.6$ and $N = 1000$. The rank-one perturbation is obtained by the least-squared $\mathbf{n}$, see Appendix D.

outlier $\lambda_j > \lambda_i$ will always destabilize the fixed point $\mathbf{x}^{(i)}$ corresponding to λ_i. Conversely, if there are no outliers with real part larger than that of λ_i, then $\mathbf{x}^{(i)}$ will be stable. Since $\lambda = 1/\langle \phi' \rangle$ implies that larger λ corresponds to larger fixed point variance Δ^0, one can say that only the largest fixed point can be stable, Such an interaction between two fixed points is illustrated in Figs. 4(b) and 4(c). The stars outside of the bulk correspond to the predicted eigenvalue γ_2 or γ_1. Comparison between the theoretical prediction (29) and numerical calculation for a sampled network shows good agreement for both fixed points. Furthermore, a simulation of the dynamics in Fig. 4(d) shows that indeed all trajectories converge to the larger fixed point $\mathbf{x}^{(1)}$ or its negative counterpart.

Finally, note that a complex outlier λ_j also destabilizes a fixed point $\mathbf{x}^{(i)}$ if the real part of λ_j is larger than that of λ_i. Complex outliers do not have corresponding fixed points, since Eq. (26) is real. An example of such a case is shown in Fig. 5. There is only one real eigenvalue larger than one, and hence only a single nontrivial fixed point $\mathbf{x}^{(3)}$. Nonetheless, the two complex outliers $\lambda_1 = \lambda_2^*$ destabilize the fixed point by virtue of Eq. (29), since the real parts are larger than the outlier corresponding to the fixed point, $\Re\lambda_1 = \Re\lambda_2 > \lambda_3$. Numerical simulations indicate that in such a case, the dynamics converge on a limit cycle.

V. RANK-TWO PERTURBATIONS

The previous section demonstrated two properties of networks with multiple nontrivial fixed points: they are highly correlated due to the confinement on the manifold $\mathcal{M}$ [Figs. 3(b)–3(d) and 4(d)], and their stability properties interact [Eq. (29)]. We asked whether the latter is a result of the former. To approach this question, we extend the model to a rank-two perturbation which allows for uncorrelated fixed points.

The rank-two connectivity structure is defined by

$$P = \mathbf{m}\mathbf{n}^T + \mathbf{u}\mathbf{v}^T. \tag{30}$$

We assume J, $\mathbf{m}$ and $\mathbf{u}$ to be drawn independently. Similar to the rank-one case, the entries of both $\mathbf{m}$ and $\mathbf{u}$ are drawn from

standard normal distributions while $\mathbf{n}$ and $\mathbf{v}$ are Gaussian and dependent on J, $\mathbf{m}$ and $\mathbf{u}$.

The outliers λ of the perturbed matrix $J + \mathbf{m}\mathbf{n}^T + \mathbf{u}\mathbf{v}^T$ are calculated similarly to the rank-one case. Applying the matrix determinant lemma twice, we arrive at an equation of quadratic form:

$$0 = \lambda^2 - \lambda \mathrm{Tr} Q_\lambda + \det(Q_\lambda). \tag{31}$$

In other words, λ is an eigenvalue of the matrix

$$Q_\lambda = \begin{bmatrix} \mathbf{n}^T M_\lambda \mathbf{m} & \mathbf{n}^T M_\lambda \mathbf{u} \\ \mathbf{v}^T M_\lambda \mathbf{m} & \mathbf{v}^T M_\lambda \mathbf{u} \end{bmatrix}, \tag{32}$$

which depends on λ through

$$M_\lambda = \left(\mathbb{1} - \frac{J}{\lambda} \right)^{-1}. \tag{33}$$

In general there are more than two solutions, but if the rank-two perturbation is uncorrelated to J, the matrix M_λ disappears in Q_λ. The solution is then in agreement with previous results [4].

Nontrivial fixed points of the network dynamics (2) with a rank-two perturbation (30) obey the equation

$$\mathbf{x} = J\boldsymbol{\phi} + \kappa_1 \mathbf{m} + \kappa_2 \mathbf{u}, \tag{34}$$

with $\kappa_1 = \mathbf{n}^T \boldsymbol{\phi}$ and $\kappa_2 = \mathbf{v}^T \boldsymbol{\phi}$. Similar to the rank-one case, we can apply the recursive insertion of the fixed point and partial integration, Eqs. (19) and (22), to compute the two-component vector $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)$. We arrive at

$$Q_\lambda \boldsymbol{\kappa} = \frac{1}{\langle \phi' \rangle} \boldsymbol{\kappa}. \tag{35}$$

This equation has two consequences: First, we find that $\lambda = 1/\langle \phi' \rangle$, because both sides are eigenvalues of Q_λ, see Eq. (31). Second, the feedback vector $\boldsymbol{\kappa}$ is the corresponding eigenvector. This gives rise to three situations.

(i) If Q_λ has two distinct eigenvalues, one of them is equal to λ. The corresponding eigenvector determines the direction of $\boldsymbol{\kappa}$.
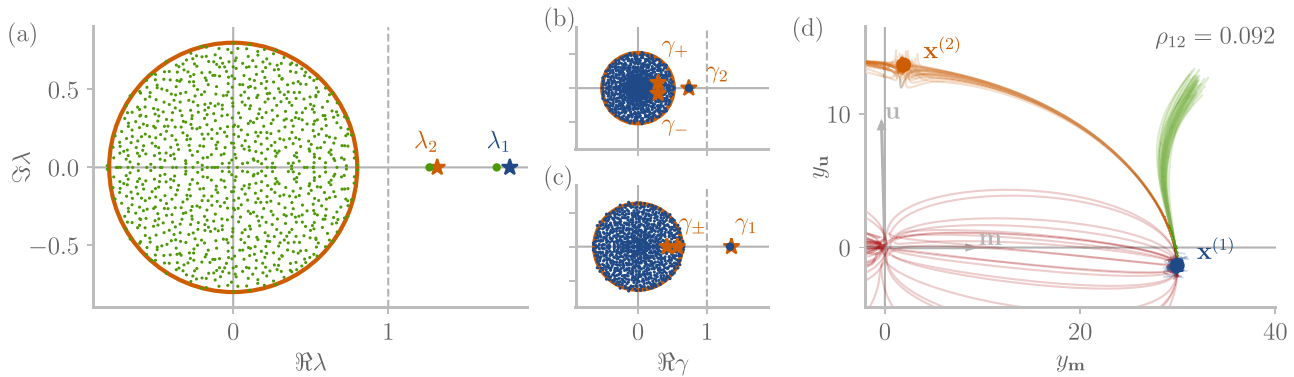
FIG. 6. Fixed points and dynamics for a rank-two perturbation with structures $\mathbf{mn}^T$ and $\mathbf{uv}^T$ drawn independently of each other as well as of J. [(a)–(c)] Spectra of the Jacobian at the origin (a) and at the two fixed points $\mathbf{x}^{(1)}$ (b) and $\mathbf{x}^{(2)}$ (c). Stars denote the predictions for infinite size networks. (d) Projection of fixed points on vectors $\mathbf{m}$ and $\mathbf{u}$, and trajectories starting around $\mathbf{x}^{(1)}$ (blue), $\mathbf{x}^{(2)}$ (orange), $\mathbf{x}^{(1)} + \mathbf{x}^{(2)}$ (green) and $\mathbf{0}$ (red). The correlation between the two fixed points is indicated by ρ_{12}. Other parameters as in Fig. 4.

In the case of degeneracy, the geometric multiplicity, that is, the number of eigenvectors, determines the situation.

(ii) If there is only one eigenvector, the direction of $\boldsymbol{\kappa}$ is determined uniquely.

(iii) If λ has two corresponding eigenvectors, any direction is a solution. The length of $\boldsymbol{\kappa}$ is determined below, Eq. (36), and we obtain a ring attractor [4]. This situation arises in the case of precise symmetry, $Q_\lambda = \lambda\mathbb{1}$.

Finally, the length of $\boldsymbol{\kappa}$ is determined by the variance $\Delta^0 = \mathbf{x}^T\mathbf{x}/N$ of the fixed point, which obeys

$$\Delta^0 = g^2\langle\phi^2\rangle + \kappa_1^2 + \kappa_2^2. \tag{36}$$

The fixed point stability is calculated based on the techniques introduced above; details can be found in Appendix H. The result is the same as that in the rank-one case: the stability eigenvalues obey the same equations as before. Namely, if the spectrum of $J + \mathbf{mn}^T + \mathbf{uv}^T$ has the outliers $\{\lambda_1, \ldots, \lambda_K\}$, there are always two stability eigenvalues $\gamma_\pm$, both with real parts smaller than one. At a fixed point $\mathbf{x}^{(i)}$, there are $K - 1$ additional outliers $\gamma_j = \lambda_j/\lambda_i$ for $j \neq i$. This implies that linear dynamics around a fixed point is completely determined by its statistics and the spectrum of the connectivity matrix: as long as the outliers are the same, the stability eigenvalues are independent of the rank of the perturbation P or its correlations to J.

This also answers our question about whether the correlation between fixed points is responsible for the strong influence on each other. The rank-two case, too, can be analyzed by replacing the feedback κ_1, κ_2 with two constant scalars. The corresponding manifold is now two-dimensional, and fixed points can be arbitrarily uncorrelated. In Fig. 6, we show an example: plotting the projection of the fixed points along the vectors $\mathbf{m}$ and $\mathbf{u}$ shows that the fixed points are almost orthogonal. Yet, the spectra at the origin and at each fixed point are identical to the corresponding rank-one case (compare with Fig. 4). The correlation between fixed points is hence not important for the mutual influence of different fixed points.

VI. DISCUSSION

Given a network with connectivity consisting of a random and a structured part, we examined the effects of correlations between the two. We found that such correlations enrich the functional repertoire of the network. This is reflected in the number of trivial fixed points and the spectrum of the connectivity matrix. We analyzed precisely which aspects of the correlations determine the fixed points and eigenvalues.

In our model, the overlaps $\theta_k = \mathbf{n}^T J^k \mathbf{n}$ quantify the correlations between random connectivity J and structured, low-rank connectivity $\mathbf{mn}^T$. For uncorrelated networks, only θ_0 is nonzero, and the spectrum of the joint connectivity matrix has only a single outlier [25,26]. We showed that in correlated networks with θ_k nonzero for higher k, multiple outliers can exist, and that with such, multiple fixed points induced by a random plus rank-one connectivity structure become possible. The correlations between random part and rank-one structure hence enrich the dynamical repertoire in contrast to networks with uncorrelated rank-one structures, which can only induce a single fixed point [4]. Note, however, that our assumption of Gaussian connectivity limits the resulting dynamics to a single stable fixed point (discussed below).

Apart from multiple fixed points, the correlated rank-one structure can also lead to a pair of complex conjugate outliers, which in turn yield oscillatory dynamics on a limit cycle. In absence of correlations, such dynamics need the perturbation to be at least of rank two [4]. Finally, we found that correlations amplify the perturbation due to the structured components: the norm of a correlated rank-one structure inducing a fixed point decreases with increasing variance of the random part, pointing towards possible benefits of large initial random connectivity.

Constraining the model to Gaussian connectivity allowed us to analytically understand the mechanisms of correlations in a nonlinear network. We established a remarkable one-to-one correspondence between the outliers of the connectivity matrix and fixed points of the nonlinear dynamics: each real outlier larger than one induces a single fixed point. Surprisingly, the stability of the fixed points is governed by a simple set of equations and also only depend on the outliers of the spectrum at the origin. Through these results, we were able to look at the system at one point in phase space (the origin) and determine its dynamics at a different part of the phase space. It remains an open question to which degree these insights

extend to non-Gaussian connectivity. Interesting other connectivity models might include sparse connectivity [30–32], different neuron types [33], or networks of binary neurons such as the Hopfield model [9].

Our approach allows us to gain mechanistic insight into the computations underlying echo state and FORCE learning models which have the same connectivity structure as our model [20,21]. Here, the readout vector **n** is trained, which leads to correlations to the random part J [3,34]. Our results on multiple fixed points and oscillations show that these correlations are crucial for the rich functional repertoire. However, constraining our theory to Gaussian connectivity limits the insights, since the learning frameworks do not have this constraint. One study analyzing such non-Gaussian rank-one connectivity in the echo state framework shows that, like in our study, each fixed point had one corresponding outlier in the connectivity matrix [3]. However, multiple stable fixed points were observed, which is in contrast to our model where the Gaussian connectivity only permits the fixed point with largest variance to be stable. It would thus be interesting to extend our model beyond the Gaussian statistics.

We pointed out a general limitation of networks with random plus rank-one connectivity: the restriction of fixed points to a one-dimensional manifold. This insight is independent of the Gaussian assumption and leads to high correlations between fixed points. Such correlations have been found to impede sequential learning of multiple fixed points [29]. An extension to rank-two structures allows for uncorrelated fixed points. Surprisingly, however, the strong influence of the largest outliers on the stability of fixed points still exists for Gaussian rank-two connectivity. Indeed, the fixed point statistics and their stability is determined solely by the spectral outliers of the connectivity matrix, independently of how these outliers were generated. Since these relations do not hold in the non-Gaussian case, [3], we conclude that the Gaussian assumption poses a severe limitation to the space of solutions.

Further in accordance with the echo state and FORCE learning frameworks [20,21], we model the correlations to be induced by one of the two vectors forming the rank-one structure. Some of the results, such as the overlaps θ_k, are symmetric under the exchange of the two vectors and should hence be unaffected. The result on the strongly increasing norm of the perturbation when placing multiple outliers, on the other hand, may depend on this assumption [28]. To which degree our results or the capabilities of trained networks are limited by this constraint is not clear.

Our choice to model the structured part as a low-rank matrix was in part motivated by the computational models discussed above. Besides these, the existence of such structures may also be inspired by a biological perspective. Any feedback loop from a high-dimensional network through an effector with a small number of degrees of freedom may be considered as a low-rank perturbation to the high-dimensional network. Similarly, feedback loops from cortex through basal ganglia have been modeled as low-rank connectivity [28]. Even without such explicit loops, networks may effectively have such structure if their connectivity is scale-free or contains hubs [35]. Finally, low-rank connectivity also appears

outside of neuroscience, for example in an evolutionary setting [36]. Whether low-rank matrices arise in general in learning networks, and to which degree such structure is correlated with the initially present connectivity are interesting future questions to be approached with the theory we developed here.

APPENDIX A: FINITE NUMBER OF SOLUTIONS FOR OUTLIERS

Equation (6) in the main text is a polynomial of infinite degree. It may thus seem that there are infinite solutions for the outlying eigenvalues, even in the case of a finite network of size N. We now show why this is not the case. Note that the equation was only valid for outliers (enabling the series expansion). Denote the eigenvalues of J by $\hat{\lambda}_i$, $1 \leqslant i \leqslant N$. If we diagonalize $J = L^T \hat{\Lambda} R$ with $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_N)$, and write $\hat{\mathbf{n}} = L\mathbf{n}$ and $\hat{\mathbf{m}} = R\mathbf{m}$, then the series becomes

$$\lambda = \sum_{k=0}^{\infty} \frac{\sum_{i=1}^{N} \hat{n}_i \hat{\lambda}_i^k \hat{m}_i}{\lambda^k} = \sum_{i=1}^{N} \frac{\hat{n}_i \hat{m}_i}{1 - \frac{\hat{\lambda}_i}{\lambda}}, \tag{A1}$$

which is a polynomial equation of degree N.

APPENDIX B: CONSTRUCTION OF THE VECTOR n

In our model, J and **m** are drawn independently from Gaussian distributions. We construct **n** from the these two quantities for a target set of overlaps $\hat{\theta}_k$. Specifically, we set

$$\mathbf{n} = \frac{1}{N} \sum_{k=0}^{\infty} \frac{\hat{\theta}_k}{g^{2k}} J^k \mathbf{m}. \tag{B1}$$

Below we show that with this definition the actual overlaps θ_k converge to the targets $\hat{\theta}_k$ with increasing network size N. Note that the scaling by $1/N$ renders the θ_k order one quantities.

We start by analyzing the uncorrelated case, for which $\hat{\theta}_k = 0$ for any $k \geqslant 1$, and

$$\mathbf{n} = \frac{\hat{\theta}_0}{N} \mathbf{m}. \tag{B2}$$

We need to show that θ_0 converges to $\hat{\theta}_0$ as $N \to \infty$. To this end, we will show that the expected value has this limit and that the variance vanishes. We calculate the scalar product

$$\theta_0 = \mathbf{n}^T \mathbf{m} = \hat{\theta}_0 \frac{\mathbf{m}^T \mathbf{m}}{N}. \tag{B3}$$

The expected value of $\mathbf{m}^T\mathbf{m}/N$ is one, so we observe that, indeed, $\mathbb{E}[\theta_0] = \hat{\theta}_0$. For the variance we have

$$
\begin{aligned}
\mathrm{var}\left(\frac{\mathbf{m}^T\mathbf{m}}{N}\right) &= \frac{1}{N^2}\sum_{ij}\mathbb{E}[m_i^2 m_j^2] - 1 \\
&= \frac{1}{N^2}\sum_{ij}\mathbb{E}[m_i^2]\mathbb{E}[m_j^2] - 1 \\
&\quad + \frac{1}{N^2}\left(\sum_i \mathbb{E}[m_i^4] - \sum_i \mathbb{E}[m_i^2]^2\right) \\
&= \mathcal{O}\left(\frac{1}{N}\right).
\end{aligned}
\tag{B4}
$$

The term in the second line vanishes, and the one in line three is of order $\mathcal{O}(1/N)$ since the fourth moment does not depend on the network size. The scalar product hence has a self-averaging quality in the sense that it converges to its expected value with deviations on the order $\mathcal{O}(1/\sqrt{N})$.

The next overlap is treated similarly:

$$
\theta_1 = \mathbf{n}^T J\mathbf{m} = \frac{\hat{\theta}_0}{N}\sum_{ij}m_i J_{ij} m_j.
\tag{B5}
$$

Here, the expected value is equal to zero because of the independence between J and $\mathbf{m}$. The variance decays with N as before:

$$
\begin{aligned}
\mathrm{var}(\theta_1) &= \left(\frac{\hat{\theta}_0}{N}\right)^2 \sum_{ijkl}\mathbb{E}[m_i J_{ij} m_j\, m_k J_{kl} m_l] \\
&= \left(\frac{\hat{\theta}_0}{N}\right)^2 \sum_{ij}\mathbb{E}[m_i^2 J_{ij}^2 m_j^2] \\
&= \left(\frac{\hat{\theta}_0}{N}\right)^2 \sum_{ij}\mathbb{E}[m_i^2 m_j^2]\frac{g^2}{N} \\
&= \theta_0^2 g^2\left(\frac{1}{N} + \frac{2}{N^2}\right) \\
&= \frac{\theta_0^2 g^2}{N} + \mathcal{O}\left(\frac{g^2}{N^2}\right).
\end{aligned}
\tag{B6}
$$

The off-diagonal terms in the first line disappear since entries with different indices are independent. Similar calculations can be done for any of the higher order overlaps θ_k, and one finds that the variance of these terms is

$$
\mathrm{var}(\theta_k) = \frac{\theta_0^2 g^{2k}}{N} + \mathcal{O}\left(\frac{g^{2k}}{N^2}\right).
\tag{B7}
$$

We now turn to the correlated case. Here, the terms $\hat{\theta}_k$ may be nonzero. We only discuss the simplest case with $\hat{\theta}_k = 0$ for any $k \geqslant 2$, since all other cases can be treated similarly. We write

$$
\mathbf{n} = \frac{1}{N}\left(\hat{\theta}_0\mathbf{m} + \frac{\hat{\theta}_1}{g^2}J\mathbf{m}\right),
\tag{B8}
$$

and calculate the zeroth overlap

$$
\theta_0 = \hat{\theta}_0\frac{\mathbf{m}^T\mathbf{m}}{N} + \frac{\hat{\theta}_1}{g^2}\frac{\cancel{\mathbf{m}^T J\mathbf{m}}}{N}.
\tag{B9}
$$

The crossed-out term self-averages to zero due to the independence between J and $\mathbf{m}$, so that $\mathbb{E}[\theta_0] = \hat{\theta}_0$.

Similar reasoning applies to the first overlap:

$$
\theta_1 = \hat{\theta}_1\frac{\cancel{\mathbf{m}^T J^T\mathbf{m}}}{N} + \frac{\hat{\theta}_1}{g^2}\frac{\mathbf{m}^T J^T J\mathbf{m}}{N}.
\tag{B10}
$$

Now, however, it is the first term that vanishes. The second one remains order $\mathcal{O}(1)$:

$$
\begin{aligned}
\mathbb{E}\left[\frac{\mathbf{m}^T J^T J\mathbf{m}}{N}\right] &= \frac{1}{N}\sum_{ijk}\mathbb{E}[m_i J_{ji} J_{jk} m_k] \\
&= \frac{1}{N}\sum_{ij}\mathbb{E}[m_i^2]\mathbb{E}[J_{ji}^2] \\
&= g^2,
\end{aligned}
\tag{B11}
$$

and hence $\mathbb{E}[\theta_1] = \hat{\theta}_1$. Similar calculations show that

$$
\mathrm{var}\left(\frac{\mathbf{m}^T J^T J\mathbf{m}}{N}\right) = \frac{2g^4}{N} + \mathcal{O}\left(\frac{g^4}{N^2}\right),
\tag{B12}
$$

where the factor 2 is of combinatorial nature, counting how many index combinations yield nonzero expectations.

The last calculations explain the scaling by $1/g^{2k}$ in the definition of $\mathbf{n}$, Eq. (B1). It also points to a general feature of the algebra of scalar products: products of the sort $\mathbf{a}(J^T)^k J^l\mathbf{b}$ for vectors $\mathbf{a}$ and $\mathbf{b}$ independent of J yield the expected value

$$
\mathbb{E}[\mathbf{a}^T(J^T)^k J^l\mathbf{b}] = \begin{cases} g^{2k}\mathbf{a}^T\mathbf{b} & \text{if}\quad k = l, \\ 0 & \text{else.} \end{cases}
\tag{B13}
$$

Applying this algebra, we see that for the construction of $\mathbf{n}$ according to Eq. (B1) one indeed obtains $\mathbb{E}[\theta_k] = \hat{\theta}_k$, valid in expectation and with variances on the order $\mathcal{O}(C_k g^{2k}/N)$, with a combinatorial factor C_k.

APPENDIX C: CONSTRUCTION OF OUTLIERS

Here we detail how to construct a rank-one perturbation $\mathbf{mn}^T$ such that the joint matrix $J + \mathbf{mn}^T$ has a set $\Lambda = \{\lambda_1, \ldots, \lambda_K\}$ of K outliers. Applying Eq. (B1) for $\mathbf{n}$, the question reduces to finding the coefficients $\hat{\theta}_k$.

The procedure of determining the $\hat{\theta}_k$ from the λ_i allows some choices. We start by choosing whether the $\hat{\theta}_k$ should form a truncated series or decay exponentially. For the truncated case with $\hat{\theta}_k = 0$ for all $k \geqslant K$, the equation follows directly from outlier equation (6):

$$
0 = \lambda^K - \sum_{k=0}^{K-1}\hat{\theta}_k\lambda^{K-1-k}.
\tag{C1}
$$

To obtain K outliers from a nontruncated series, one can write the $\hat{\theta}_k$ as sums of exponentially decaying terms

$$
\hat{\theta}_k = \sum_{\alpha=1}^{K}a_\alpha b_\alpha^k,
\tag{C2}
$$

with coefficients a_α and bases b_α. Evaluating the geometric series leads to the polynomial equation

$$
1 = \sum_{\alpha=1}^{K}\frac{a_\alpha}{\lambda - b_\alpha}.
\tag{C3}
$$

Either choice hence yields a polynomial of degree K, the roots of which need to be the target outliers λ_i. The coefficients of this polynomial can hence be obtained from a comparison with the coefficients of the polynomial $p(\lambda) = \prod_{i=1}^{K}(\lambda - \lambda_i)$. For example, in the case of a truncated series, the coefficients $\hat{\theta}_k$ are determined by

$$\hat{\theta}_k = (-1)^k \sum_{\pi \in S_{\binom{K}{k+1}}} \left(\prod_{i \in \pi} \lambda_i \right), \tag{C4}$$

where $S_{\binom{K}{k}}$ denotes the possible choices of k different indices from the K available ones. Note that the indices for θ_k run from zero to $K - 1$, whereas those of λ_i run from 1 to K.

In the case of exponentially decaying series, one can derive a similar equation from Eq. (C3). However, there are $2K$ parameters a_α and b_α–there is the freedom to choose K of them, as long as $b_\alpha < g$.

APPENDIX D: LEAST SQUARE VECTOR n

Here we introduce another method of constructing multiple outliers. Assume that J, $\mathbf{m}$ are given and we want to find the least square vector $\mathbf{n}$ such that we have the set of outliers $\{\lambda_1, \ldots \lambda_K\}$. We apply the matrix determinant lemma, Eq. (4), and obtain

$$\mathbf{n}^T (\lambda_\alpha \mathbb{1} - J)^{-1} \mathbf{m} = 1 \quad \forall \quad \alpha \in \{1, \ldots, K\}. \tag{D1}$$

This can be read as an underconstrained linear system $A\mathbf{n} = \mathbf{1}$, with the vector of ones $\mathbf{1}$ and the matrix $A \in \mathbb{C}^{K \times N}$ defined by its rows

$$A_\alpha^T = \frac{1}{\lambda_\alpha} \mathbf{w}_{\lambda_\alpha}, \tag{D2}$$

with $\mathbf{w}_\lambda = (\mathbb{1} - \frac{J}{\lambda})^{-1} \mathbf{m}$. Since $K < N$, the matrix A is singular. The least square solution to the system is given by

$$\mathbf{n} = A^+ \mathbf{1}, \tag{D3}$$

with the pseudoinverse of A denoted by $A^+ = A^T (AA^T)^{-1} \in \mathbb{C}^{N \times K}$.

We express the vector $\mathbf{w}_\lambda$ by the series expansion

$$\mathbf{w}_\lambda = \sum_k \left(\frac{J}{\lambda} \right)^k \mathbf{m}, \tag{D4}$$

and insert this into the term AA^T. Applying the algebra (B13) developed above yields

$$\begin{aligned}
(AA^T)_{\alpha\beta} &= \sum_{i=1}^{N} A_{\alpha i} A_{\beta i} \\
&= \frac{1}{\lambda_\alpha \lambda_\beta} \sum_{k,l} \mathbf{m}^T \left(\frac{J^T}{\lambda_\alpha} \right)^k \left(\frac{J}{\lambda_\beta} \right)^l \mathbf{m} \\
&= \frac{1}{\lambda_\alpha \lambda_\beta} \sum_k \left(\frac{g^2}{\lambda_\alpha \lambda_\beta} \right)^k \mathbf{m}^T \mathbf{m} \\
&= \frac{N}{\lambda_\alpha \lambda_\beta - g^2}.
\end{aligned} \tag{D5}$$

Inserting this into Eq. (D3), we can finally write

$$\mathbf{n} = \frac{1}{N} \sum_{\alpha=1}^{K} a_\alpha \mathbf{w}_{\lambda_\alpha}, \tag{D6}$$

where the coefficients a_α are determined by solving the linear equation

$$1 = \sum_{\beta=1}^{K} \frac{a_\beta}{\lambda_\alpha - g^2/\lambda_\beta} \quad \forall \quad \alpha \in \{1, \ldots, K\}. \tag{D7}$$

Connecting to the above schemes of constructing $\mathbf{n}$ with deliberate overlaps $\theta_k = \mathbf{n}^T J^k \mathbf{m}$, we compute these for the least square solution found here. We find that overlaps decay exponentially as in Eq. (C2), namely,

$$\theta_k = \sum_{\alpha=1}^{K} a_\alpha \left(\frac{g^2}{\lambda_\alpha} \right)^k. \tag{D8}$$

We observed numerically that for a large number of outliers K the rank-one perturbation became the dominant term in the matrix $J + \mathbf{m}\mathbf{n}^T$. To understand this, we look at its Frobenius norm, given by

$$||\mathbf{m}\mathbf{n}^T||^2 = \text{Tr}((\mathbf{m}\mathbf{n}^T)^T \mathbf{m}\mathbf{n}^T) = N\mathbf{n}^T\mathbf{n}. \tag{D9}$$

The squared norm of $\mathbf{n}$ can be obtained from the pseudoinverse defined above:

$$\mathbf{n}^T\mathbf{n} = \mathbf{1}^T (A^+)^T A^+ \mathbf{1} = \mathbf{1}^T (AA^T)^{-1} \mathbf{1}. \tag{D10}$$

To arrive at a general expression for this quantity, we calculated Eq. (D5) explicitly for the cases $K = 1$ and 2 and performed thorough numerical checks for larger K (see Fig. 2). The resulting equation is

$$\mathbf{n}^T\mathbf{n} = \frac{g^2}{N} \left(\prod_{\alpha=1}^{K} \frac{\lambda_\alpha^2}{g^2} - 1 \right). \tag{D11}$$

For an increasing number of outliers, the offset by minus one becomes negligible, and we arrive at the result stated in the main text, Eq. (12).

APPENDIX E: MEAN-FIELD THEORY WITH CORRELATIONS

We analyze fixed points of the form

$$\mathbf{x} = J\boldsymbol{\phi} + \kappa \mathbf{m} + \mathbf{I}. \tag{E1}$$

We extend the setting in the main text by allowing for a constant input vector $\mathbf{I}$. Like the other vectors, we assume $\mathbf{I}$ to be Gaussian and uncorrelated to J.

We want to compute $\kappa = \mathbf{n}^T\phi(\mathbf{x})$. The mean-field assumption is that $\mathbf{x}$ is a Gaussian variable. That is, its entries x_i are drawn from a normal distribution with zero mean and variance Δ^0. The variance Δ^0 is determined self-consistently below. Since the entries of the vector follow the same statistics, we look at a representative x_i and drop the index i:

$$x = \sqrt{\Delta^0} z_x, \tag{E2}$$

with the standard Gaussian random variable $z_x \sim \mathcal{N}(0, 1)$. By the model assumption, the vector $\mathbf{n}$ is also a Gaussian. One

can express **n** explicitly in relation to **x** by defining

$$n = \frac{1}{N}(\sigma_n \sqrt{1-\rho}\, z_n + \sigma_n \rho\, z_x), \tag{E3}$$

with a second, independent standard Gaussian random variable z_n. The parameters σ_n^2 and ρ encode the variance of **n** and its correlation to **x**. The self-averaging quality of the scalar product allows us to write

$$\mathbf{n}^T \mathbf{x} = N\,\mathbb{E}[nx] = \sqrt{\Delta^0}\sigma_n \rho. \tag{E4}$$

Computing κ can then be achieved by exchanging between the scalar product $\mathbf{n}^T \boldsymbol{\phi}$ and the corresponding Gaussian integral:

$$\begin{aligned}
\kappa &= \mathbf{n}^T \boldsymbol{\phi} \\
&= \int \mathcal{D}z_x \int \mathcal{D}z_n (\sigma_n \sqrt{1-\rho}\, z_n + \sigma_n \rho\, z_x)\phi(\sqrt{\Delta^0}\, z_x) \\
&= \sqrt{\Delta^0}\sigma_n \rho \int \mathcal{D}z_x \phi'(\sqrt{\Delta^0}\, z_x) \\
&= \mathbf{n}^T \mathbf{x}\,\langle \phi' \rangle,
\end{aligned} \tag{E5}$$

where $\mathcal{D}z$ is the standard Gaussian measure. In the second line, the term $\sigma_n \sqrt{1-\rho}z_n$ vanishes with the integration over z_n. For the second summand, the integrating over z_n evaluates to 1. The step from second to third line involves partial integration: for some function f and the Gaussian variable $z \sim \mathcal{N}(0,1)$, we have $\int \mathcal{D}y\, z f(z) = \int \mathcal{D}z\, f'(z)$. This is also known as Stein's lemma [4]. Finally, the angled brackets $\langle \cdot \rangle$ indicate the average over the fixed point statistics,

$$\langle \phi' \rangle = \int \mathcal{D}z\, \phi'(\sqrt{\Delta^0}z). \tag{E6}$$

Such explicit representations of pairs of correlated Gaussian variables have been applied before [3,4]. Below, however, we encounter a multitude of such Gaussian vectors, so such a framework would become increasingly cumbersome. We thus introduce a new formalism which allows us to model arbitrary many Gaussian vectors. Let **a** and **b** be two such vectors of interest. Like before, we are interested in the statistics of a representative entry, namely a and b. We define these as

$$a = B_a \cdot X, \quad b = B_b \cdot X. \tag{E7}$$

Here, $B_a, B_b \in \mathbb{R}^K$ are a set of real-valued coefficients and X is a K-dimensional standard normal Gaussian variable with mutually independent entries $X_\alpha \sim \mathcal{N}(0,1)$. The K-dimensional dot product is defined as $B_a \cdot X = \sum_{\alpha=1}^{K}(B_a)_\alpha X_\alpha$. Any additional vectors **c**, **d**, ... are added by defining corresponding coefficients $B_c, B_d, \ldots$ One just needs to choose the dimension K of the embedding to be sufficiently large.

Since scalar products in the N-dimensional physical space are self-averaging, we can write

$$\frac{1}{N}\mathbf{a}^T \mathbf{b} = \frac{1}{N}\sum_{i=1}^{N} a_i b_i = B_a \cdot B_b. \tag{E8}$$

In line with the previous sections, the equality sign is only valid in the limit $N \to \infty$, and the variance decays like $1/N$. Ultimately, we are interested in such scalar products. This allows us to use the coefficients B_a merely as placeholders inside calculations, without ever defining their actual values.

With this notation we return to the computation of κ:

$$\begin{aligned}
\kappa &= \mathbf{n}^T \boldsymbol{\phi} \\
&= N \int \mathcal{D}X\,(B_n \cdot X)\,\phi(B_x \cdot X) \\
&= NB_n \cdot B_x \int \mathcal{D}X\, \phi'(B_x \cdot X) \\
&= \mathbf{n}^T \mathbf{x}\,\langle \phi' \rangle,
\end{aligned} \tag{E9}$$

where $\mathcal{D}X$ is the now standard Gaussian measure in K dimensions. The third line is obtained using partial integration as before. In the last line of Eq. (E9) above, we inserted the definition of coefficients from above, Eq. (E8), for the scalar products.

The advantage of the new formalism is that it allows us to continue the calculation. We insert the fixed point $\mathbf{x} = J\boldsymbol{\phi} + \kappa\mathbf{m} + \mathbf{I}$. In the case of structure vectors drawn independently from J, the term $\mathbf{n}^T J\boldsymbol{\phi}$ vanishes and one recovers the known result $\kappa = \langle \phi' \rangle\, \mathbf{n}^T(\kappa\mathbf{m} + \mathbf{I})$ [4]. For the general case, we go on calculating $\mathbf{n}^T J\boldsymbol{\phi}$. The scalar product allows to pull the random matrix to the left side, and hence

$$\begin{aligned}
\mathbf{n}^T J\boldsymbol{\phi} &= N \int \mathcal{D}X (B_{J^T n} \cdot X)\,\phi(B_x \cdot X) \\
&= \langle \phi' \rangle\, \mathbf{n}^T J\mathbf{x}.
\end{aligned} \tag{E10}$$

Recursively applying this strategy, we arrive at

$$\begin{aligned}
\kappa &= \langle \phi' \rangle\, \mathbf{n}^T \sum_k \left(\langle \phi' \rangle J\right)^k (\kappa\mathbf{m} + \mathbf{I}) \\
&= \langle \phi' \rangle\, \mathbf{n}^T M(\kappa\mathbf{m} + \mathbf{I}),
\end{aligned} \tag{E11}$$

with

$$M = (\mathbb{1} - \langle \phi' \rangle J)^{-1}. \tag{E12}$$

For a driven network with nonzero $\mathbf{n}^T M\mathbf{I}$, one can compute κ by resorting:

$$\kappa = \frac{\langle \phi' \rangle \mathbf{n}^T M\mathbf{I}}{1 - \langle \phi' \rangle \mathbf{n}^T M\mathbf{m}}. \tag{E13}$$

The scalar $\langle \phi' \rangle$, Eq. (25), is a function of the fixed point variance $\Delta^0 = \mathbf{x}^T \mathbf{x}/N$. Due to the independence of **m** and **I** from J, the variance obeys the same equation as in previous studies [4]:

$$\Delta^0 = g^2 \langle \phi^2 \rangle + (\kappa\mathbf{m} + \mathbf{I})^T(\kappa\mathbf{m} + \mathbf{I})/N, \tag{E14}$$

with $\langle \phi^2 \rangle = \int \mathcal{D}z\, \phi^2(\sqrt{\Delta^0}z)$. The coupled nonlinear Eqs. (25), (E13), and (E14) can be solved numerically if the overlaps $\theta_k = \mathbf{n}^T J^k \mathbf{m}$ and $\mathbf{n}^T J^k \mathbf{I}$, which respectively enter $\mathbf{n}^T M\mathbf{m}$ and $\mathbf{n}^T M\mathbf{I}$, are known.

In the case of no input, $\mathbf{I} = 0$, κ is not directly determined. Instead, $\langle \phi' \rangle$ is given directly by the outliers via Eq. (26), as discussed in the main text. The mean-field equations can be closed by numerically solving Eq. (E6) for Δ^0. The corresponding κ is determined up to the sign by

$$\Delta^0 = g^2 \langle \phi^2 \rangle + \kappa^2. \tag{E15}$$

APPENDIX F: STABILITY EIGENVALUES

The matrix $JR' - \mathbb{1}\gamma$ with the diagonal matrix $R'_{ij} = \delta_{ij}\phi'_i$ is invertible as long as $\gamma \in \mathbb{C}$ is not within the spectrum of JR'. We apply the matrix determinant lemma to compute the characteristic polynomial,

$$\det(JR' - \mathbb{1}\gamma) = (1 + \mathbf{n}^T R'(JR' - \mathbb{1}\gamma)^{-1}\mathbf{m})\det(JR' - \mathbb{1}\gamma).$$
(F1)

The first bracket has to vanish, so that we arrive at

$$\gamma = \mathbf{n}^T R'\left(\mathbb{1} - \frac{JR'}{\gamma}\right)^{-1}\mathbf{m} = \sum_k \frac{\mathbf{n}^T R'(JR')^k\mathbf{m}}{\gamma^k}, \quad (F2)$$

for a nonzero γ.

We calculate the terms $\mathbf{n}^T R'(JR')^k\mathbf{m}$ applying the Gaussian mean-field theory introduced above. For brevity we use induction. The hypothesis to be proven is

$$\mathbf{n}^T R'(JR')^k\mathbf{m} = \mathbf{n}^T J^k\mathbf{m}\langle\phi'\rangle^{k+1} + \frac{\mathbf{x}^T\mathbf{m}}{N}\langle\phi'''\rangle$$
$$\times \sum_{l=0}^k q^l \mathbf{n}^T(\langle\phi'\rangle J)^{k-l}\mathbf{x}, \quad (F3)$$

where $\mathbf{x}$ denotes the fixed point and we define

$$q = g^2\langle\phi''\phi + \phi'^2\rangle. \quad (F4)$$

We start the induction by calculating the Gaussian integral

$$\mathbf{n}^T R'\mathbf{m} = N\int \mathcal{D}X \,(B_n \cdot X)\,\phi'(B_x \cdot X)\,B_m \cdot X$$
$$= \mathbf{n}^T\mathbf{m}\,\langle\phi'\rangle + \mathbf{n}^T\mathbf{x}\,\frac{\mathbf{x}^T\mathbf{m}}{N}\,\langle\phi'''\rangle. \quad (F5)$$

We used partial differentiation twice, which yields the third derivative of the nonlinearity. The steps above did not depend the specific vectors $\mathbf{m}$ and $\mathbf{n}$ so we will apply the same step below without explicitly mentioning the Gaussian integrals. The entire deviation furthermore does not depend on the vector $\mathbf{n}$. We make use of this independence in the induction step where we assume the hypothesis (F3) to be true also after replacing $\mathbf{n}^T$ with $\mathbf{n}^T J$. This term is identified by square brackets in the below calculation:

$$\mathbf{n}^T R'(JR')^k\mathbf{m}$$
$$= \mathbf{n}^T\left(\langle\phi'\rangle + \frac{\mathbf{x}\mathbf{x}^T}{N}\langle\phi'''\rangle\right)(JR')^k\mathbf{m}$$
$$= [\mathbf{n}^T J]R'(JR')^{k-1}\mathbf{m}\,\langle\phi'\rangle + \mathbf{n}^T\mathbf{x}\frac{1}{N}\mathbf{x}^T(JR')^k\mathbf{m}\,\langle\phi'''\rangle$$
$$= [\mathbf{n}^T J]J^{k-1}\mathbf{m}\,\langle\phi'\rangle^k\langle\phi'\rangle + \mathbf{n}^T\mathbf{x}\frac{\mathbf{x}^T\mathbf{m}}{N}\langle\phi'''\rangle$$
$$\times \sum_{l=0}^{k-1} q^l[\mathbf{n}^T J](\langle\phi'\rangle J)^{k-1-l}\mathbf{x}\langle\phi'\rangle + \mathbf{n}^T\mathbf{x}\frac{\mathbf{x}^T\mathbf{m}}{N} q^k\,\langle\phi'''\rangle.$$
(F6)

The first step is to replace $\mathbf{m}$ with $(JR')^k\mathbf{m}$ in Eq. (F5). The step from second to third line involves the induction hypothesis for $k-1$, and including the last term in the sum

completes the induction. The last term needs to be calculated separately. We show that

$$\mathbf{x}^T(JR')^k\mathbf{m} = \mathbf{x}^T\mathbf{m}\,q^k \quad (F7)$$

in a separate induction. The start is trivial. For the induction step at $k \geq 1$, we insert the fixed point equation $\mathbf{x} = J\boldsymbol{\phi} + \kappa\mathbf{m} + \mathbf{I}$.

$$\mathbf{x}^T(JR')^k\mathbf{m} = \boldsymbol{\phi}^T J^T(JR')^k\mathbf{m} + \underline{(\kappa\mathbf{m} + \mathbf{I})^T(JR')^k\mathbf{m}}$$
$$= \boldsymbol{\phi}^T J^T J[R'(JR')^{k-1}\mathbf{m}]$$
$$= g^2\boldsymbol{\phi}^T[R'(JR')^{k-1}\mathbf{m}]$$
$$= q\,\mathbf{x}^T(JR')^{k-1}\mathbf{m}. \quad (F8)$$

Inserting the induction hypothesis for $k-1$ proves the statement. A few comments on the steps of the calculation. (1) The crossed-out term in the first line vanishes because $\mathbf{m}$ and $\mathbf{I}$ are drawn independently of J. Showing this formally is a matter of applying the same techniques as introduced above recursively. (2) The step from line two to three involves the vector algebra (B13) introduced above, according to which $\mathbf{a}^T J^T J\mathbf{b} = g^2\mathbf{a}^T\mathbf{b}$ for two vectors $\mathbf{a}$, $\mathbf{b}$ independent of J. (3) The fourth line is obtained by applying partial integration. For an arbitrary Gaussian vector $\mathbf{a}$, we have

$$\boldsymbol{\phi}^T R'\mathbf{a} = N\int \mathcal{D}X \,\phi(B_x \cdot X)\,\phi'(B_x \cdot X)\,B_a \cdot X$$
$$= \mathbf{x}^T\mathbf{a}\langle\phi''\phi + \phi'^2\rangle. \quad (F9)$$

We go back to the eigenvalue equation (F2) and insert Eq. (F3), which yields

$$\gamma = \sum_k \frac{\mathbf{n}^T R'(JR')^k\mathbf{m}}{\gamma^k}$$
$$= \langle\phi'\rangle \sum_k \left(\frac{\langle\phi'\rangle}{\gamma}\right)^k\mathbf{n}^T J^k\mathbf{m}$$
$$+ \frac{\mathbf{x}^T\mathbf{m}}{N}\langle\phi'''\rangle \sum_k \left(\frac{q}{\gamma}\right)^k \sum_{l=0}^k \left(\frac{\langle\phi'\rangle}{q}\right)^l\mathbf{n}^T J^l\mathbf{x}. \quad (F10)$$

Note that we swap the order of summation in the second sum. This sum evaluates to

$$\sum_k \left(\frac{q}{\gamma}\right)^k \sum_{l=0}^k \left(\frac{\langle\phi'\rangle}{q}J\right)^l$$
$$= \sum_k \left(\frac{q}{\gamma}\right)^k\left(\mathbb{1} - \frac{\langle\phi'\rangle}{q}J\right)^{-1}\left[\mathbb{1} - \left(\frac{\langle\phi'\rangle}{q}J\right)^{k+1}\right]$$
$$= \left(\mathbb{1} - \frac{\langle\phi'\rangle}{q}J\right)^{-1}\left[\frac{1}{1 - \frac{q}{\gamma}} - \frac{\langle\phi'\rangle}{q}J\left(\mathbb{1} - \frac{\langle\phi'\rangle}{\gamma}J\right)^{-1}\right]$$
$$= \left(\mathbb{1} - \frac{\langle\phi'\rangle}{q}J\right)^{-1}\left[\frac{1}{1 - \frac{q}{\gamma}}\left(\mathbb{1} - \frac{\langle\phi'\rangle}{\gamma}J\right)^{-1}\left(\mathbb{1} - \frac{\langle\phi'\rangle}{q}J\right)\right]$$
$$= \frac{1}{1 - \frac{q}{\gamma}}\left(\mathbb{1} - \frac{\langle\phi'\rangle}{\gamma}J\right)^{-1}. \quad (F11)$$

Inserting this into Eq. (F2) for γ, we obtain

$$\gamma = \mathbf{n}^T \left(\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J \right)^{-1} \left[\langle \phi' \rangle \mathbf{m} + \frac{\langle \phi''' \rangle}{1 - \frac{q}{\gamma}} \frac{\mathbf{m}^T \mathbf{x}}{N} \mathbf{x} \right]. \quad \text{(F12)}$$

One can further simplify this expression by inserting the fixed point $\mathbf{x} = J\phi(\mathbf{x}) + \kappa \mathbf{m} + \mathbf{I}$ and evaluating the Gaussian statistics. In particular, we have

$$\mathbf{a}^T \mathbf{x} = \mathbf{a}^T (\mathbb{1} - \langle \phi' \rangle J)^{-1} (\kappa \mathbf{m} + \mathbf{I}) \quad \text{(F13)}$$

for any Gaussian vector $\mathbf{a}$. In particular,

$$\mathbf{n}^T \left(\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J \right)^{-1} \mathbf{x} = \mathbf{n}^T \left(\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J \right)^{-1}$$
$$\times (\mathbb{1} - \langle \phi' \rangle J)^{-1} (\kappa \mathbf{m} + \mathbf{I}). \quad \text{(F14)}$$

The product of the two inverses can be conveniently split. For any matrix A and a scalar a, completion of the denominator yields

$$\frac{1}{1-a}[(\mathbb{1} - A)^{-1} - a(\mathbb{1} - aA)^{-1}]$$
$$= \frac{1}{1-a}[(\mathbb{1} - aA) - a(\mathbb{1} - A)](\mathbb{1} - aA)^{-1}(\mathbb{1} - A)^{-1}$$
$$= (\mathbb{1} - aA)^{-1}(\mathbb{1} - A)^{-1}. \quad \text{(F15)}$$

Inserting this into Eq. (F14), we arrive at

$$\gamma = \langle \phi' \rangle \mathbf{n}^T \left(\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J \right)^{-1} \mathbf{m} + \frac{\langle \phi''' \rangle}{\left(1 - \frac{q}{\gamma} \right)\left(1 - \frac{1}{\gamma} \right)} \frac{\mathbf{m}^T \mathbf{x}}{N}$$
$$\times \left[\frac{\kappa}{\langle \phi' \rangle} - \frac{1}{\gamma} \mathbf{n}^T \left(\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J \right)^{-1} (\kappa \mathbf{m} + \mathbf{I}) \right]. \quad \text{(F16)}$$

We made use of the Eq. (E13) constraining κ, namely

$$\frac{\kappa}{\langle \phi' \rangle} = \mathbf{n}^T (\mathbb{1} - \langle \phi' \rangle J)^{-1} (\kappa \mathbf{m} + \mathbf{I}). \quad \text{(F17)}$$

Re-sorting terms finally results in

$$\left[\frac{\gamma}{\langle \phi' \rangle} - \mathbf{n}^T \left(\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J \right)^{-1} \mathbf{m} \right]$$
$$= \frac{\langle \phi''' \rangle \kappa (\kappa + \mathbf{m}^T \mathbf{I}/N)\gamma}{\langle \phi' \rangle (\gamma - q)(\gamma - 1)}$$
$$\times \left[\frac{\gamma}{\langle \phi' \rangle} - \mathbf{n}^T \left(\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J \right)^{-1} \left(\mathbf{m} + \frac{1}{\kappa} \mathbf{I} \right) \right]. \quad \text{(F18)}$$

Note that we also inserted $\mathbf{m}^T \mathbf{x}/N = \kappa + \mathbf{m}^T \mathbf{I}/N$. Without knowledge about the interaction between $\mathbf{n}$ and $\mathbf{I}$, we cannot further simplify since the term $\mathbf{n}^T (\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J)^{-1} \mathbf{m}$ without $\mathbf{I}$ and the other term including it, $\mathbf{n}^T (\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J)^{-1} (\kappa \mathbf{m} + \mathbf{I})$, appear in two different parts of the theory: the first one is a property of the matrix, related to the outlier equation $\lambda = \mathbf{n}^T (\mathbb{1} - \frac{1}{\lambda} J)^{-1} \mathbf{m}$, the second one determines κ.

APPENDIX G: STABILITY EIGENVALUES FOR THE AUTONOMOUS NETWORK

In the autonomous case, $\mathbf{I} = \mathbf{0}$, the square brackets in Eq. (F18) become identical, and the equation splits into a quadratic part and one of degree $K - 1$. By applying the identity $\lambda = 1/\langle \phi' \rangle$, we arrive at

$$0 = \left[\lambda \gamma - \mathbf{n}^T \left(\mathbb{1} - \frac{J}{\lambda \gamma} \right)^{-1} \mathbf{m} \right] \left[\frac{\langle \phi''' \rangle \kappa^2 \lambda \gamma}{(\gamma - q)(\gamma - 1)} - 1 \right]. \quad \text{(G1)}$$

The second bracket is squared in γ and exhibits the roots

$$\gamma_{\pm} = \frac{c}{2} \pm \sqrt{\left(\frac{c}{2} \right)^2 - q}, \quad \text{(G2)}$$

with

$$c = 1 + q + \kappa^2 \frac{\langle \phi''' \rangle}{\langle \phi' \rangle}, \quad \text{(G3)}$$

and $q = g^2 \langle \phi'' \phi + \phi'^2 \rangle$ as defined above, Eq. (F4). We observe that $\gamma_{\pm}$ is entirely defined by the fixed point statistics. One can even reduce the problem to two parameters, for example the outlier λ and the network parameter g which quantifies the strength of the random connectivity. This allows to thoroughly scan the numerical values of $\gamma_{\pm}$. The results can be observed in Fig. 7. The first observation is that both γ_+ and γ_- are always smaller than one. They hence do not destabilize the fixed point. Additionally, we can compare $\gamma_{\pm}$ to the radius of the bulk, which is given by $r = g\sqrt{\langle \phi'^2 \rangle}$ [4]. This shows that γ_- is always within the bulk and hence not observable numerically.

The roots of the first bracket in Eq. (G1) are identified by comparing once again with Eq. (5) which defines the outliers λ of the spectrum at the origin. In fact, the equation is identical, but now the variable is $\lambda \gamma$. Here, λ is the outlier corresponding to the fixed point under consideration. We hence need to fulfill $\lambda \gamma = \lambda'$ for some λ' in the set of outliers. Accordingly, the solutions are given by Eq. (29) in the main text.

The case $\lambda = \lambda'$ and hence $\gamma = 1$ was omitted in the discussion above. However, one observes from Eq. (F12) after insertion of $\lambda = 1/\langle \phi' \rangle$ that $\gamma = 1$ would necessitate $\mathbf{n}^T (\mathbb{1} - J/\lambda)^{-2} \mathbf{m} = 0$. According to Eq. (5), the eigenvalues λ are the roots of the function $f(\lambda) = \mathbf{n}^T (\mathbb{1} \lambda - J)^{-1} \mathbf{m} - 1$. The function has the derivative $df/d\lambda = -\mathbf{n}^T (\mathbb{1} \lambda - J)^{-2} \mathbf{m}$, which only vanishes if the roots have multiplicity larger than 1. Thus $\gamma = 1$ is only a solution if the corresponding λ has algebraic multiplicity larger than 1.

APPENDIX H: STABILITY FOR RANK-TWO PERTURBATION

The stability of a fixed point in the case of a rank-two perturbation can be evaluated similarly to the rank-one case. One simply applies the matrix determinant lemma once more on the stability matrix. Let $\lambda = 1/\langle \phi' \rangle$ be the outlier corresponding to the fixed point under consideration. All mean-field quantities will hence implicitly depend on λ, even though we omit this dependency in the notation. The quadratic equation
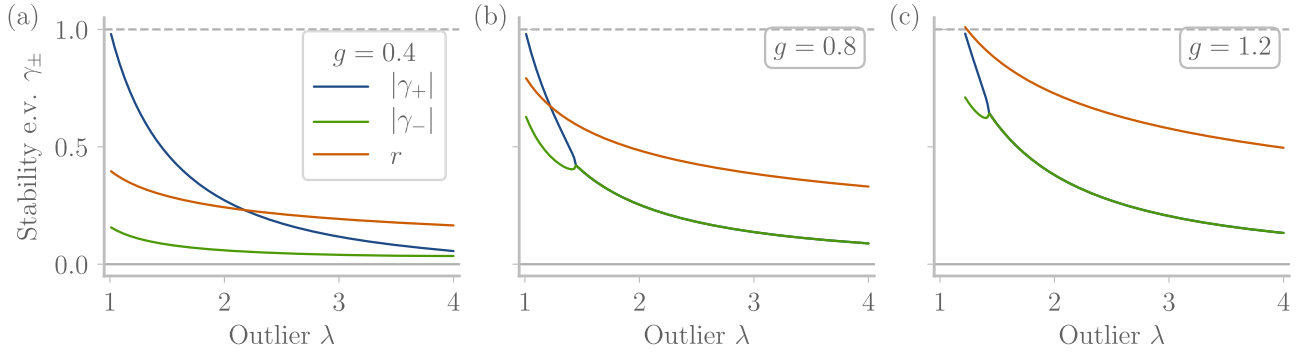
FIG. 7. Stability eigenvalues $\gamma_{\pm}$ depend on the outlier λ corresponding to a fixed point. The $\gamma_{\pm}$ are solutions to the mean-field equations, fully determined by the eigenvalue λ and random strength g. (a)–(c) correspond to $g = (0.4, 0.8, 1.2)$, respectively. Orange lines: radius of the bulk; any eigenvalues with smaller magnitude are not observable, and hence not numerically testable for finite size networks. Where the absolute values of $\gamma_{\pm}$ coincide, the two form a pair of complex conjugates. Note that for $g > 1$, the minimal λ to stabilize the chaotic activity is larger than one.

for the stability eigenvalues γ reads

$$0 = \gamma^2 - \gamma \text{Tr}\tilde{Q}_\gamma + \det(\tilde{Q}_\gamma), \qquad \text{(H1)}$$

with the mean-field form of the stability matrix

$$\tilde{Q}_\gamma = \begin{bmatrix} \mathbf{n}^T \tilde{M}_\gamma \mathbf{m} & \mathbf{n}^T \tilde{M}_\gamma \mathbf{u} \\ \mathbf{v}^T \tilde{M}_\gamma \mathbf{m} & \mathbf{v}^T \tilde{M}_\gamma \mathbf{u} \end{bmatrix}, \qquad \text{(H2)}$$

and $\tilde{M}_\gamma = R'(\mathbb{1} - JR'/\gamma)^{-1}$. As above, Eq. (F12), we have

$$\mathbf{a}^T \tilde{M}_\gamma \mathbf{b} = \mathbf{a}^T \left(\mathbb{1} - \frac{\langle \phi' \rangle}{\gamma} J \right)^{-1} \left(\langle \phi' \rangle \mathbf{b} + \frac{\langle \phi''' \rangle}{1 - \frac{q}{\gamma}} \mathbf{b}^T \mathbf{x} \, \mathbf{x} \right)$$

$$= \frac{1}{\lambda} \mathbf{a}^T M_{\lambda\gamma} \mathbf{b} + \frac{\langle \phi''' \rangle \gamma}{(\gamma - q)(\gamma - 1)} \mathbf{b}^T (\kappa_1 \mathbf{m} + \kappa_2 \mathbf{u})$$

$$\times \mathbf{a}^T (\gamma M_\lambda - M_{\lambda\gamma})(\kappa_1 \mathbf{m} + \kappa_2 \mathbf{u}), \qquad \text{(H3)}$$

for two vectors $\mathbf{a}$ and $\mathbf{b}$ and $M_\lambda = (\mathbb{1} - J/\lambda)^{-1}$ as before, Eq. (33). The second line is valid in case of an autonomous

fixed point, cf. Eq. (G1). Evaluating the terms appearing in $\tilde{Q}_\gamma$, we arrive at

$$\tilde{Q}_\gamma = \frac{1}{\lambda}[Q_{\lambda\gamma} + A \, \Delta Q \, \boldsymbol{\kappa}\boldsymbol{\kappa}^t]. \qquad \text{(H4)}$$

We abbreviated $A = \frac{\langle \phi''' \rangle \lambda \gamma}{(\gamma - q)(\gamma - 1)}$ and $\Delta Q = \gamma Q_\lambda - Q_{\lambda\gamma}$. The trace and determinant are then conveniently evaluated as

$$\det(\tilde{Q}_\gamma) = \frac{\det(Q_{\lambda\gamma})}{\lambda^2}(1 + A\boldsymbol{\kappa}^t(Q_{\lambda\gamma})^{-1}\Delta Q\boldsymbol{\kappa}), \qquad \text{(H5)}$$

$$\text{Tr}(\tilde{Q}_\gamma) = \frac{1}{\lambda}(\text{Tr}(Q_{\lambda\gamma}) + A\boldsymbol{\kappa}^t \Delta Q\boldsymbol{\kappa}). \qquad \text{(H6)}$$

Inserting these expressions into the stability eigenvalue equation (H1) and recalling that $\boldsymbol{\kappa}$ is an eigenvector of Q_λ, Eq. (35), we finally arrive at

$$0 = [(\lambda\gamma)^2 - \lambda\gamma\text{Tr}(Q_{\lambda\gamma}) + \det(Q_{\lambda\gamma})][1 - A\boldsymbol{\kappa}^t\boldsymbol{\kappa}]. \qquad \text{(H7)}$$

We hence arrive at the same solutions as in the case of a rank-one perturbation, Eqs. (29) and (G2).

[1] M. Rigotti, O. Barak, M. R. Warden, X.-J. Wang, N. D. Daw, E. K. Miller, and S. Fusi, The importance of mixed selectivity in complex cognitive tasks, Nature (London) **497**, 585 (2013).

[2] G. R. Yang, M. R. Joglekar, H. F. Song, W. T. Newsome, and X.-J. Wang, Task representations in neural networks trained to perform many cognitive tasks, Nat. Neurosci. **22**, 297 (2019).

[3] A. Rivkind and O. Barak, Local Dynamics in Trained Recurrent Neural Networks, Phys. Rev. Lett. **118**, 258101 (2017).

[4] F. Mastrogiuseppe and S. Ostojic, Linking connectivity, dynamics, and computations in low-rank recurrent neural networks, Neuron **99**, 609 (2018).

[5] B. Tirozzi and M. Tsodyks, Chaos in highly diluted neural networks, Europhys. Lett. **14**, 727 (1991).

[6] Y. Roudi and P. E. Latham, A balanced memory network, PLoS Comput. Biol. **3**, e141 (2007).

[7] Y. Ahmadian, F. Fumarola, and K. D. Miller, Properties of networks with partially structured and partially random connectivity, Phys. Rev. E **91**, 012820 (2015).

[8] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Theory of orientation tuning in visual cortex, Proc. Natl. Acad. Sci. USA **92**, 3844 (1995).

[9] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[10] X.-J. Wang, Probabilistic decision making by slow reverberation in cortical circuits, Neuron **36**, 955 (2002).

[11] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Spin-glass models of neural networks, Phys. Rev. A **32**, 1007 (1985).

[12] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, Chaos in Random Neural Networks, Phys. Rev. Lett. **61**, 259 (1988).

[13] K. Rajan, L. F. Abbott, and H. Sompolinsky, Stimulus-dependent suppression of chaos in recurrent neural networks, Phys. Rev. E **82**, 011903 (2010).

[14] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, J. Comput. Neurosci. **8**, 183 (2000).

[15] G. Wainrib and J. Touboul, Topological and Dynamical Complexity of Random Neural Networks, Phys. Rev. Lett. **110**, 118101 (2013).

[16] C. Van Vreeswijk and H. Sompolinsky, Chaos in neuronal networks with balanced excitatory and inhibitory activity, Science **274**, 1724 (1996).

[17] C. Huang, D. A. Ruff, R. Pyle, R. Rosenbaum, M. R. Cohen, and B. Doiron, Circuit models of low-dimensional shared variability in cortical networks, Neuron **101**, 337 (2019).

[18] A. Renart, R. Moreno-Bote, X.-J. Wang, and N. Parga, Mean-driven and fluctuation-driven persistent activity in recurrent networks, Neural Comput. **19**, 1 (2007).

[19] W. Maass, T. Natschläger, and H. Markram, Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Comput. **14**, 2531 (2002).

[20] H. Jaeger and H. Haas, Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication, Science **304**, 78 (2004).

[21] D. Sussillo and L. F. Abbott, Generating coherent patterns of activity from chaotic neural networks, Neuron **63**, 544 (2009).

[22] P. Enel, E. Procyk, R. Quilodran, and P. F. Dominey, Reservoir computing properties of neural dynamics in prefrontal cortex, PLoS Comput. Biol. **12**, e1004967 (2016).

[23] O. Barak, D. Sussillo, R. Romo, M. Tsodyks, and L. F. Abbott, From fixed points to chaos: Three models of delayed discrimination, Prog. Neurobiol. **103**, 214 (2013).

[24] J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, J. Math. Phys. **6**, 440 (1965).

[25] K. Rajan and L. F. Abbott, Eigenvalue Spectra of Random Matrices for Neural Networks, Phys. Rev. Lett. **97**, 188104 (2006).

[26] T. Tao, Outliers in the spectrum of iid matrices with bounded rank perturbations, Probab. Theory Relat. Fields **155**, 231 (2013).

[27] D. A. Harville, *Matrix Algebra from a Statistician's Perspective* (Taylor & Francis Group, 1998).

[28] L. Logiaco, L. F. Abbott, and S. Escola, A model of flexible motor sequencing through thalamic control of cortical dynamics, bioRxiv (2019), doi:10.1101/2019.12.17.880153.

[29] C. Beer and O. Barak, One step back, two steps forward: Interference and learning in recurrent neural networks, Neural Comput. **31**, 1985 (2019).

[30] I. Neri and F. L. Metz, Spectra of Sparse Non-Hermitian Random Matrices: An Analytical Solution, Phys. Rev. Lett. **109**, 030602 (2012).

[31] I. Neri and F. L. Metz, Eigenvalue Outliers of Non-Hermitian Random Matrices with a Local Tree Structure, Phys. Rev. Lett. **117**, 224101 (2016).

[32] F. L. Metz, I. Neri, and T. Rogers, Spectral theory of sparse non-Hermitian random matrices, J. Phys. A: Math. Theor. **52**, 434003 (2019).

[33] J. Aljadeff, M. Stern, and T. Sharpee, Transition to Chaos in Random Networks with Cell-Type-Specific Connectivity, Phys. Rev. Lett. **114**, 088101 (2015).

[34] F. Mastrogiuseppe and S. Ostojic, A geometrical analysis of global stability in trained feedback networks, Neural Comput. **31**, 1139 (2019).

[35] A. Rivkind, H. Schreier, N. Brenner, and O. Barak, Scale free topology as an effective feedback system, bioRxiv (2019), doi:10.1101/696575.

[36] C. Furusawa and K. Kaneko, Formation of dominant mode by evolution in biological systems, Phys. Rev. E **97**, 042410 (2018).