Commun. Theor. Phys. **77** (2025) 035601 (17pp) https://doi.org/10.1088/1572-9494/ad8126

An optimization-based equilibrium measure describing fixed points of non-equilibrium dynamics: application to the edge of chaos

Junbin Qiu[1] **and Haiping Huang**[1,2]

[1] PMI Lab, School of Physics, Sun Yat-sen University, Guangzhou 510275, China
[2] Guangdong Provincial Key Laboratory of Magnetoelectric Physics and Devices, Sun Yat-sen University, Guangzhou 510275, China

E-mail: huanghp7@mail.sysu.edu.cn

Received 27 July 2024, revised 28 September 2024
Accepted for publication 29 September 2024
Published 4 December 2024

Abstract

Understanding neural dynamics is a central topic in machine learning, non-linear physics, and neuroscience. However, the dynamics are non-linear, stochastic and particularly non-gradient, i.e., the driving force cannot be written as the gradient of a potential. These features make analytic studies very challenging. The common tool is the path integral approach or dynamical mean-field theory. Still, the drawback is that one has to solve the integro-differential or dynamical mean-field equations, which is computationally expensive and has no closed-form solutions in general. From the associated Fokker–Planck equation, the steady-state solution is generally unknown. Here, we treat searching for the fixed points as an optimization problem, and construct an approximate potential related to the speed of the dynamics, and find that searching for the ground state of this potential is equivalent to running approximate stochastic gradient dynamics or Langevin dynamics. Only in the zero temperature limit, can the distribution of the original fixed points be achieved. The resultant stationary state of the dynamics exactly follows the canonical Boltzmann measure. Within this framework, the quenched disorder intrinsic in the neural networks can be averaged out by applying the replica method, which leads naturally to order parameters for the non-equilibrium steady states. Our theory reproduces the well-known result of edge-of-chaos. Furthermore, the order parameters characterizing the continuous transition are derived, and the order parameters are explained as fluctuations and responses of the steady states. Our method thus opens the door to analytically studying the fixed-point landscape of the deterministic or stochastic high dimensional dynamics.

Keywords: high-dimensional chaos, phase transitions, neural networks, order parameters, statistical physics

(Some figures may appear in colour only in the online journal)

1. Introduction

Non-equilibrium dynamics is a central topic of statistical physics [1–3], from which general fluctuation-dissipation relations can be established [4–6], and a path integral approach was developed to derive the relationship between the correlation and response function of the dynamics [7–9]. One benchmark for testing or deriving theoretical hypotheses are the Langevin dynamics with conservative (expressed as the gradient of a potential) or nonconservative forces, including Ising spin glass dynamics [10], steady-state thermodynamics [11], and neural dynamics with partially symmetric couplings [12]. These studies always resort to dynamical mean-field theory (DMFT) [13, 14], where all possible trajectory paths are integrated to give a physical action, from which the correlation and response functions are derived as an optimization of the action [15]. In a general context, the dynamics would be very complicated, e.g., non-

gradient and highly coupled, which makes the DMFT very complicated, and it remains unknown how to capture the steady-state properties of these dynamics.

It is commonly observed that the speed of the dynamics determines whether the system enters a steady-state and thus can be identified [16]. In this paper, we propose a new angle that only the low-speed or zero-speed region of the phase space is considered. In this sense, one can write down an energy cost function to optimize for satisfying this speed constraint. This optimization can be formulated under the stochastic gradient dynamics, where a temperature-like parameter is introduced. This temperature plays the same role as in the canonical Boltzmann measure. In fact, in the zero temperature and stationary limit, the gradient dynamics under this approximate (or quasi) potential are self-consistent with the original general dynamics. Therefore, if the ground state of the energy cost is concentrated, one can capture the steady-state properties of the original phase space. This has an additional benefit that powerful equilibrium concepts such as the Boltzmann measure, phase transitions, and order parameters can be leveraged [17].

We verify our framework in recurrent neural networks with asymmetric couplings, where the deterministic dynamics yield an order-to-chaos transition [18, 19]. We are able to handle the quenched disorder due to the coupling statistics with the help of the replica method, a powerful trick in the spin glass theory [20]. Furthermore, two types of order parameters are naturally derived; one describes the network activity, while the other describes the response property of the dynamic system. The former order parameter exhibits a continuous phase transition from a null to positive value as the network parameters are tuned, while the latter one shows a peak at the exact location of the transition, thereby offering rich physical information about the steady-state of the complicated non-gradient dynamics. Our framework also makes it possible to analytically study the quasi-potential landscape of general dynamics, using the large-deviation principle and landscape analysis in equilibrium statistical mechanics [17, 21], which is left for a forthcoming work [22].

2. Model

We consider a recurrent neural network (RNN) composed of N interacting neurons, whose architecture is shown in figure 1(a), illustrated as a simple example of $N=3$. The network activity is described by N real-valued synaptic currents ($\boldsymbol{x}$). In the following, we use the bold symbol to indicate a vector or a matrix. The connection strength J_{ij} measures the asymmetric impact from neuron j to neuron i. Note that in our model, two neurons do not have to interact with each other equally, i.e. $J_{ij} \neq J_{ji}$. In addition, $J_{ii}=0$. All neurons are fully connected. The synaptic currents are transferred to firing rates through the current-to-rate transfer function $\phi(\cdot)$. Each neuron integrates neighboring activity via the incoming asymmetric synapses, and therefore the dynamics of the current are expressed as the following first-order non-linear differential equation for all i [18, 23]

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_i + \sum_{j=1}^{N} J_{ij}\phi(x_j), \tag{1}$$

where we consider unit time scale. Unless otherwise stated, $\phi = \tanh$. In this case, the transferred neural activity may become negative, while for biological reality, the non-negative function form should be chosen. However, our following analysis is still applicable to this type of non-negative transfer function.

In real neural circuits, J_{ij} may not be identically independently distributed (i.i.d.). We therefore consider the correlated case whose statistics are specified below

$$\langle (J_{ij})^2 \rangle = \frac{g^2}{N}, \tag{2}$$

$$\langle J_{ij} J_{ji} \rangle = \frac{g^2}{N}\gamma, \tag{3}$$

where $g>0$ is called the gain parameter, which controls the variance of connections, and $\gamma \in [-1, 1]$ controls the degree of connection symmetry. The special cases of $\gamma=1$ and $\gamma=-1$ correspond to symmetric and anti-symmetric neural networks, respectively, while $\gamma=0$ corresponds to fully asymmetric networks [18].

The above dynamics model exhibits an order-to-chaos transition [19]. The average (over the coupling realizations) number of equilibria explodes as the synaptic gain exceeds a critical value [24]. The critical gain value can be derived from the rightmost boundary of the asymptotic eigenvalue spectrum of $\boldsymbol{J}$ [25]. When $g(1+\gamma)<1$, the network stays in the trivial fixed-point phase, where all trajectories flow to the zero-activity stable state. Once $g(1+\gamma)>1$, the network activity enters a chaotic phase, where an infinitesimal separation of two trajectories would be eventually amplified, and individual neurons exhibit slow to fast time scales as the network parameters go deep into the chaotic phase. We illustrate the phase diagram in figure 1(b) and show the representative dynamic behaviors of these two phases in figure 1(c).

Intuitively, the phase space is partitioned into different regions of different speeds ($\|\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}\|^2$). We are interested in the zero-speed or low-speed regions [16]. Focusing on the speed landscape, we design the following approximate potential (quasi-potential) of the dynamics,

$$E(\boldsymbol{x}) = \frac{1}{2}\sum_i \left(-x_i + \sum_{j=1}^{N} J_{ij}\phi(x_j) \right)^2 + \eta \|\boldsymbol{x}\|^2, \tag{4}$$

where the first term is the kinetic energy, the second term is a regularization term constraining the ℓ_2 norm of the network activity. η plays the role of tuning the relative strength of the regularization to the kinetic energy. Here, we transform the studies of non-equilibrium steady states of non-gradient dynamics into an optimization of searching for the low-speed or zero-speed regions of the phase space, which corresponds to local or global minima of the above quasi-potential. This potential is equivalent to Hamiltonian in equilibrium statistical physics.

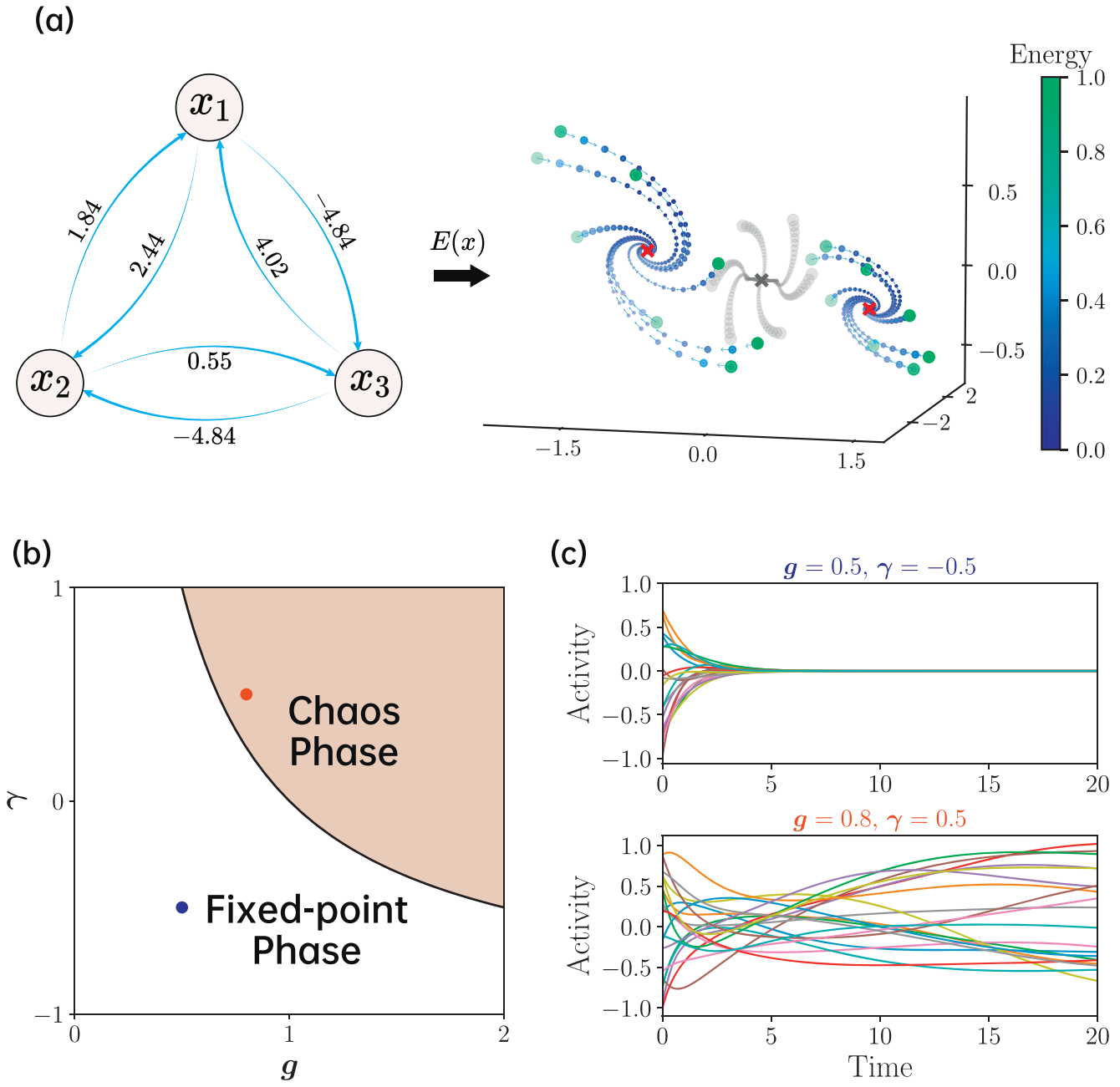

Figure 1. The key idea of our quasi-potential method and the phase diagram of the considered RNN model. (a) A three-neuron recurrent neural network is shown in the left panel. Neurons are bidirectionally connected with the strength displayed near the arrows (synaptic directions). Both connections in opposite directions are correlated once $\gamma \neq 0$. Here, we simulate the RNN dynamics [equation (1)], whose phase space displays two fixed points ($g = 1.2$, $\gamma = 0.1$). In the right panel, we show the trajectories flowing to these fixed points, where the color and size of the dots represent the energy of the corresponding position in the phase space, and the arrows represent the direction and the ℓ_2 norm of the velocity (the length of the arrow). We also show an example of the trivial fixed-point for comparison ($g = 0.8$, $\gamma = 0.1$), where nearby trajectories are shown in the light grey color. The cross symbols denote these corresponding fixed points. (b) The phase diagram as a function of gain parameter g and symmetric degree γ is divided into two regions: null fixed-point and chaotic phases, separated by an analytic transition line $g(1+\gamma) = 1$ [19]. (c) We show the representative behavior of the dynamics [$\phi(x)$] in the two phases, corresponding to two colored points in the phase diagram of (b).

According to equation (4), we can write down the following stochastic gradient dynamics (or Langevin dynamics)

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\nabla_x E(\boldsymbol{x}) + \sqrt{2T}\,\boldsymbol{\epsilon}, \tag{5}$$

where $\boldsymbol{\epsilon}(t)$ is a time-dependent white noise, whose statistics are given by $\langle \epsilon_i(t) \rangle = 0$, $\langle \epsilon_i(t)\epsilon_j(t') \rangle = \delta_{ij}\delta(t-t')$. The temperature T is used to tune the energy level, playing the same role as in the equilibrium Boltzmann measure. We can write the gradient (force) in equation (5) in the component-wise, i.e., $-\frac{\partial E(\boldsymbol{x})}{\partial x_i} \equiv F_i$ where

$$F_i \equiv -x_i + h_i - \phi'(x_i) \sum_{j:j\neq i} J_{ji}(h_j - x_j), \tag{6}$$

where $h_i \equiv \sum_{j;j\neq i} J_{ij}\phi(x_j)$, and the last term explains the feedback effects due to neuron i whose activity yields impacts on its neighbors. From this expansion, we can see that only in the stationary limit ($T \to 0$ as well), do the above Langevin dynamics yield the same state space as that of the original dynamics (e.g., stable or unstable fixed points). In this sense, the energy defined in equation (4) is not a Lyapunov function, and even the steady-state potential whose analytic form for the current out-of-equilibrium dynamics with asymmetric couplings is unknown [26], although the driving force is argued to be decomposed into the gradient of a potential (not the Lyapunov function as well) related to the steady-state distribution and a divergent-free curl flux [27]. It is still challenging to get the exact form of the potential for this decomposition. In contrast, we start from an optimization angle that concentrates on the low-speed region of the phase space and formulate the steady-state behavior as the canonical ensemble of stationary fixed points.

It is well-known that the steady-state of the above Langevin dynamics has an equilibrium Boltzmann measure [1],

$$P(\boldsymbol{x}) = \frac{1}{Z}\mathrm{e}^{-\beta E(\boldsymbol{x})}, \tag{7}$$

where the partition function Z depends on particular realizations of $\boldsymbol{J}$, and $\beta = 1/T$. An illustration of $E(\boldsymbol{x})$ is given in figure 1(a). Our main focus is to compute the free energy of this canonical ensemble and derive the fixed-point properties of the dynamics model. The theoretical predictions can be confirmed by simulating the discretized Langevin dynamics, assuming the small discretization step as a learning rate, akin to a Monte Carlo simulation. We remark that this equilibrium analysis reproduces the continuous nature of the dynamics transition, more insights and the possibility of exploring the speed landscape will be discussed in the remaining sections.

3. Replica method

To derive the quenched disorder average of the free energy, we apply the replica method, which is popular in studying the theory of spin glasses [20] and statistical mechanics of neural networks [17]. We first write down the following replicated partition function

$$\begin{aligned} Z^n &= \int \mathrm{d}\boldsymbol{x} \exp\left[-\beta\left(\frac{1}{2}\sum_{ia}\left(-x_i^a + \sum_j J_{ij}\phi(x_j^a)\right)^2 + \eta\sum_a \|\boldsymbol{x}^a\|^2\right)\right] \\ &= \int \mathrm{d}\boldsymbol{x}\mathrm{D}\hat{\boldsymbol{x}} \exp\left[\mathrm{i}\sqrt{\beta}\sum_{ia}\hat{x}_i^a\left(-x_i^a + \sum_j J_{ij}\phi(x_j^a)\right) - \beta\eta\sum_a\|\boldsymbol{x}^a\|^2\right], \end{aligned} \tag{8}$$

where to arrive at the second equality, the Hubbard–Stratonovich transformation is used [17], and i, a denote the neuron and replica index, respectively. The neuron index runs from 1 to N, while the replica index runs from 1 to n. We write a shorthand $\mathrm{D}\hat{\boldsymbol{x}} \equiv \prod_{i=1}^N \prod_{a=1}^n \mathrm{D}\hat{x}_i^a$ where $\mathrm{D}\hat{x} \equiv \mathrm{e}^{-\frac{1}{2}\hat{x}^2}\mathrm{d}\hat{x}/\sqrt{2\pi}$ being the Gaussian measure, and we call $\hat{x}$ as a response field [15]. We also denote $\mathrm{d}\boldsymbol{x}$ as a shorthand for $\prod_{i=1}^N \prod_{a=1}^n \mathrm{d}x_i^a$, where $x_i^a \in \mathbb{R}$. Therefore, $-\beta f = \lim_{n\to 0}\frac{\ln\langle Z^n\rangle}{Nn}$ [20], where $\langle\cdot\rangle$ denotes the disorder average over $\boldsymbol{J}$. The neural dynamics are much faster than the coupling dynamics, in which we assume a stationary limit of fixed $\boldsymbol{J}$, and thus the quenched disorder average is required. But if both dynamics evolve in similar time scales, an enlarged state space may be considered, and we leave this issue to future works.

In the course of replica analysis (see details in appendix A), two order parameters are naturally introduced as follows,

$$Q^{ab} = \frac{1}{N}\sum_i \phi(x_i^a)\phi(x_i^b), \tag{9}$$

$$R^{ab} = \frac{1}{N}\sum_i \hat{x}_i^a\phi(x_i^b), \tag{10}$$

where $\hat{x}_i^a$ acts as a response field, whose name is inspired by the analysis using DMFT [15] and explained below. Briefly, this response field emerges due to the linearization of the quadratic terms in the Hamiltonian [see equation (4) and equation (8)]. As the first level of approximation, we assume the replica symmetric ansätz to order parameters $Q^{ab} = q\delta_{ab} + Q(1-\delta_{ab})$ and $R^{ab} = r\delta_{ab} + R(1-\delta_{ab})$, whose physical meaning is explained in more details in appendix A. Under this approximation, the free energy is given by

$$\begin{aligned} -\beta f = &\frac{1}{2}Q\hat{Q} - q\hat{q} + R\hat{R} - r\hat{r} - \ln\sigma + \frac{1}{2}\beta g^2\gamma(r^2 - R^2) \\ &+ \int (\mathrm{D}u\mathrm{D}v)\ln I, \end{aligned} \tag{11}$$

where $\sigma \equiv \sqrt{1 + g^2\beta(q-Q)}$, and $\hat{Q}$, $\hat{q}$, $\hat{R}$ and $\hat{r}$ are conjugated order parameters. The integral $I \equiv \int \mathrm{d}x \mathrm{e}^{\mathcal{H}(x)}$, where

$$\begin{aligned} \mathcal{H}(x) \equiv &-\beta\eta x^2 + \frac{1}{2}(2\hat{q} - \hat{Q})\phi^2(x) \\ &+ \left(\sqrt{\frac{g^2\beta Q\hat{Q} - \hat{R}^2}{g^2\beta Q}}u + \frac{\hat{R}}{g\sqrt{\beta Q}}v\right)\phi(x) \\ &- \frac{1}{2\sigma^2}(g\sqrt{\beta Q}v + (\hat{r} - \hat{R})\phi(x) - \sqrt{\beta}x)^2. \end{aligned} \tag{12}$$

In the large N limit, we can use the saddle-point approximation (also called the Laplace method) to derive the closed equations the order parameters must obey. Therefore, by setting the derivatives of the replica free energy with respect to the order parameters zero, we obtain the following compact self-consistent equations (see appendix B for details)

$$q = [\langle\phi^2\rangle], \tag{13a}$$

$$Q = [\langle\phi\rangle^2], \tag{13b}$$

$$\hat{q} = -\frac{gk}{2} + \frac{g^2k^2Q}{2} + \frac{gk}{2\sigma^2}(\hat{r} - \hat{R})f(1, 1, -2)[\langle\phi^2\rangle] + \frac{gk}{\sigma^2}(\hat{r} - \hat{R})f(0, -1, 1)[\langle\phi\rangle^2] + \frac{k^2}{2}(1 - 2gkQ)[\langle x^2\rangle] + \frac{gk\sqrt{\beta}}{\sigma^2}f(0, 1, -2)[\langle x\rangle\langle\phi\rangle] + \frac{gk\sqrt{\beta}}{\sigma^2}f(-1, 0, 2)[\langle x\phi\rangle] + gk^3Q[\langle x\rangle^2], \tag{13c}$$

$$\hat{Q} = g^2k^2Q + \frac{2gk}{\sigma^2}(\hat{r} - \hat{R})f(0, 1, -1)[\langle\phi^2\rangle] + \frac{gk}{\sigma^2}(\hat{r} - \hat{R})f(1, -3, 2)[\langle\phi\rangle^2] - 2gk^3Q[\langle x^2\rangle] - \frac{2gk\sqrt{\beta}}{\sigma^2}f(1, -2, 2)[\langle x\rangle\langle\phi\rangle] + k^2(1 + 2gkQ)[\langle x\rangle^2] + \frac{2gk\sqrt{\beta}}{\sigma^2}f(0, -1, 2)[\langle x\phi\rangle], \tag{13d}$$

$$r = -\frac{1}{\sigma^2}f(1, 0, -1)[\langle\phi^2\rangle] + \frac{1}{\sigma^2}f(0, 1, -1)[\langle\phi\rangle^2] + \frac{\sqrt{\beta}}{\sigma^2}(1 - gkQ)[\langle\phi x\rangle] + \frac{\sqrt{\beta}}{\sigma^2}gkQ[\langle\phi\rangle\langle x\rangle], \tag{13e}$$

$$R = -\frac{1}{\sigma^2}f(0, 1, -1)[\langle\phi^2\rangle] - \frac{1}{\sigma^2}f(1, -2, 1)[\langle\phi\rangle^2] - \frac{\sqrt{\beta}}{\sigma^2}gkQ[\langle x\phi\rangle] + \frac{\sqrt{\beta}}{\sigma^2}(1 + gkQ)[\langle x\rangle\langle\phi\rangle], \tag{13f}$$

$$\hat{r} = \beta g^2\gamma r, \tag{13g}$$

$$\hat{R} = \beta g^2\gamma R, \tag{13h}$$

where $k \equiv \frac{g\beta}{\sigma^2}$ and $f(a, b, c) \equiv a\hat{r} + b\hat{R} + cgkQ(\hat{r} - \hat{R})$. In addition, two kinds of averages are specified. One is the disorder average $[\cdot] \equiv \int \mathrm{D}u\mathrm{D}v\cdot$, and the other is the thermal average $\langle\cdot\rangle \equiv \frac{\int \mathrm{d}x e^{\mathcal{H}(x)}\cdot}{\int \mathrm{d}x e^{\mathcal{H}(x)}}$. $\mathcal{H}(x)$ plays the role of the effective Hamiltonian of our model. We emphasize that this thermal average is only defined in this saddle-point equation, or other quantities that can be expressed as a function of order parameters; otherwise, $\langle\cdot\rangle$ should be referred to as the average over the coupling statistics. When $\gamma = 0$ (i.i.d. scenario), $\hat{r}$, $\hat{R}$, and $f(a, b, c)$ vanish, the above saddle point equation can be greatly simplified (see appendix C).

We finally remark that the above two kinds of order parameters arising from the quenched disorder average have clear physical meanings. First, $q - Q$ characterizes the activity fluctuation in the neural population, which can be directly measured in neural circuits. Second, R^{ab} can be interpreted as the response function. More precisely, the steady-state receives a perturbation of synaptic currents (e.g., denoted by I_a, where a can be assumed as a state index [20]), and the linear response $\frac{\partial\langle\phi_b\rangle}{\partial I_a}$ characterizes how the system reacts to the small perturbation. ϕ_b is the mean-field activity of the state b. Note that the average $\langle\cdot\rangle$ is defined under equation (8), which leads to $R^{ab} = \langle\hat{x}^a\phi_b\rangle$. Therefore we explain R^{ab} as a linear response in physics. We shall support this picture with theoretical arguments and numerical simulations in the next section.

4. Results and discussion

4.1. Equilibrium replica analysis reveals the continuous nature of the chaos transition

The zero temperature limit selects the ground state of the Boltzmann measure [equation (7)], which captures the statistics of the steady-state solutions of the original non-gradient dynamics [equation (1)]. An explicit form of the Lyapunov function that is monotonically decreasing over time can only be identified provided that the neural coupling is symmetric, and only in this special case, the Lyapunov function reduces to an Ising-model like Hamiltonian, as encountered in standard Hopfield networks [28]. We thus remark that the quasi-potential defined in equation (4) is not the physical potential of the original dynamics, yet capturing only the steady-state properties of the dynamics, e.g., the low- (or zero-) speed landscape. In this sense, not all characteristics of non-gradient dynamics are included in our approximate potential description, e.g., the transient behavior or the temporal evolution of auto-correlation even in the time translation invariant regime is not captured. However, the statistics of fixed-point solutions can be well captured by the quasi-potential approach.

The average energy is an indicator of whether the RNN dynamics are trapped in fixed points. In other words, steady states are achievable provided that the average energy can be optimized to the ground state zero value. Without loss of generality, we consider the independent setting $\gamma = 0$ and set $\eta = 0$. As shown in figure 2, with increasing value of β, the energy decreases to 0, implying that the ground state can be achieved in a very large β. When $\beta = 10000$, the energy becomes less than 10^{-4}, suggesting that it is reasonable to set $\beta = 10000$ for the zero temperature limit considered in our numerical solutions of the self-consistent equation [see equation (13*a*)]. We also show the ℓ_2 norm of the network activity in figure 2. The norm displays different magnitudes before and after the onset of chaos. We also plot a comparison between the gradient Langevin dynamics [see equation (5)] at a very low temperature and the original dynamics in figure 3. The approximate gradient dynamics are able to reach the fixed points of the original dynamics in the presence of relatively small g and remain there; in other words, the chaos is made stable. In the original dynamics, because of instability (at least one eigenvalue (real part) of the state-dependent Jacobian is positive), those fixed points outline the realm of chaotic dynamics. It is intuitively expected that as the temperature for the Langevin dynamics goes to zero, both dynamics share the same set of fixed points [see also the rationale we explained earlier in the previous paragraph and that below equation (6)]. We leave a systematic exploration of this point in a future work.

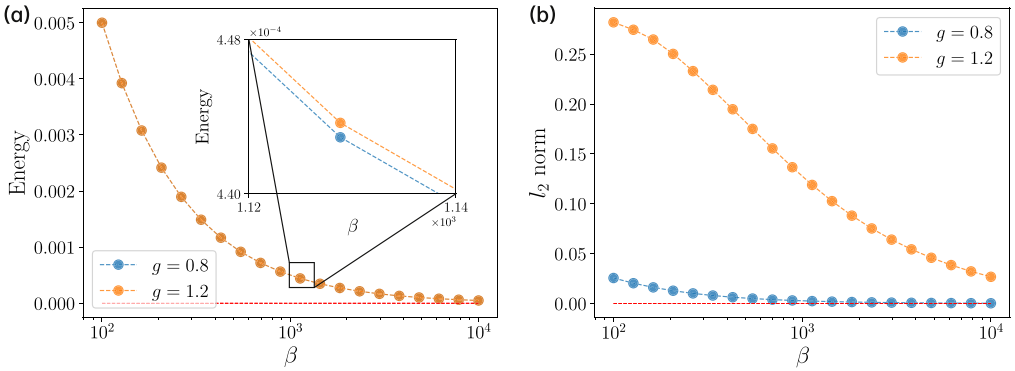

Figure 2. The average energy (a) and ℓ_2 norm (b) of network activity versus the inverse temperature β in the i.i.d. scenario $\gamma = 0$ with $\eta = 0$. The orange and the blue curves represent the setting $g = 0.8$ and $g = 1.2$, respectively. The red dashed line denotes the ground state energy. Two lines of $g = 0.8$ and $g = 1.2$ are nearly indistinguishable.

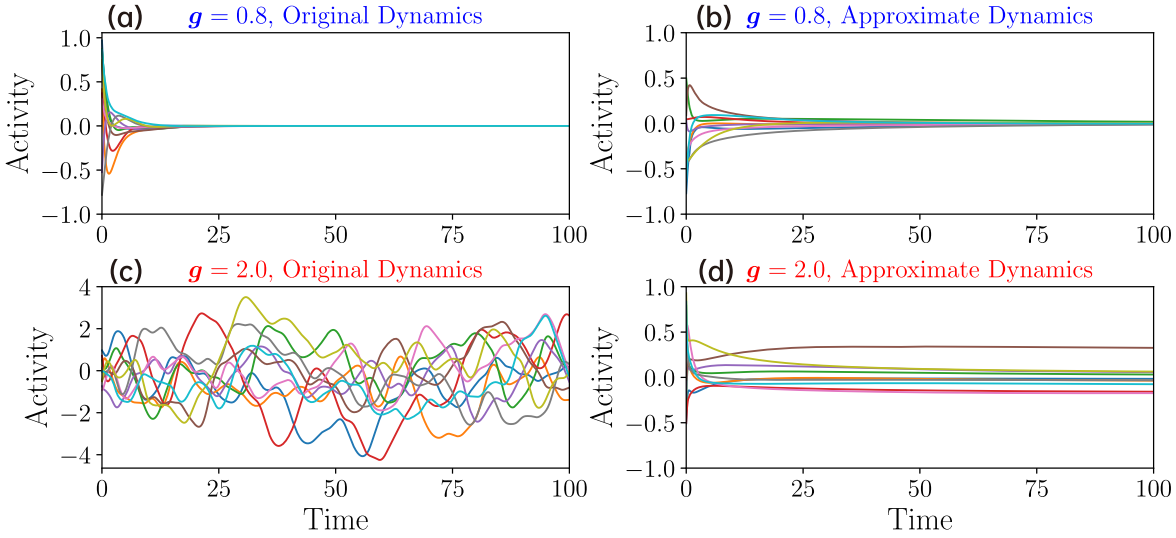

Figure 3. Dynamics comparison between the original non-gradient one and approximate gradient one at zero temperature. Representative trajectories of $\boldsymbol{x}$ are shown. Note that when $g > 1$ and $\gamma = 0$, the approximate dynamics flow to fixed points, which outline the realm of the original chaotic dynamics.

With a finite temperature, the low-speed solutions of the original dynamics could also be studied. For the remaining analysis, we focus on a very large β (e.g., $\beta = 10^4$) and set $\eta = 0$ unless otherwise stated. Varying the control parameters g and γ, we plot the profiles of all order parameters (q, Q, r, R) in figure 4.

Our replica analysis reveals exactly the order-to-chaos transition [19]. When $g < \frac{1}{1+\gamma}$, the zero fixed-point uniquely dominates the phase space, resulting in vanishing q and Q, as expected from the physical definition of the order parameters. Once $g > \frac{1}{1+\gamma}$, q and Q increase rapidly but continually from zero, indicating the emergence of self-sustained collective activity. This coincides with the known picture of exponentially growing topological complexity [24] and the positive maximal Lyapunov exponent [17, 18, 29]. It would be interesting to count the stationary points of our potential, rather than using the hard-to-compute Kac–Rice formula [24], which we leave for future thorough work [22]. Our method thus provides a relatively simple way to access the dynamics landscape in the steady-state limit. We remark that the steady-state distribution of the network activity for the original non-gradient dynamics does not have an analytic form to date.

4.2. Response order parameter for non-equilibrium steady states

Let us now look at the associated order parameters involving the response field $\hat{x}$, which emerges due to the linearization of the quadratic terms in the Hamiltonian (see details in appendix A). This response field is similar to that used to enforce the dynamics equation in a field-theoretic formula of the stochastic dynamics [15]. However, in our current setting, the response field is time-independent, as only steady-state properties are focused on in this work. Therefore, we argue that R and r bear the same physical meaning with the steady-state integrated response function in the DMFT language [15]. Let us explain in detail. Based on the derivation of DMFT in our recent work [15], we have the one-dimensional mean-field dynamics equation:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -x(t) + \gamma g^2 \int_0^\infty R(t, t')\phi(t')\mathrm{d}t' + \omega(t), \quad (14)$$

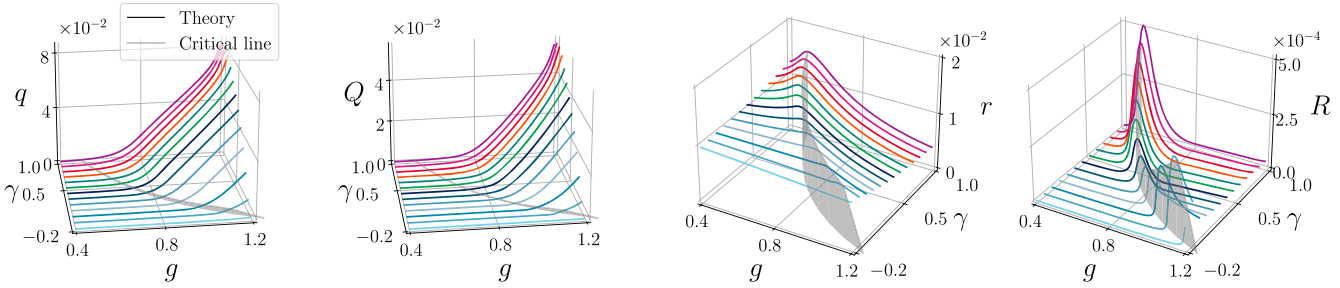

Figure 4. Order parameter profile as a function of gain parameters g and symmetric degree γ. g is varied from 0.4 to 1.2, and γ is varied from -0.2 to 1.0 with an interval of 0.1. The colors of curves correspond to the different γ values. The grey vertical plane specifies the location of the critical line $g(\gamma + 1) = 1$ [19].

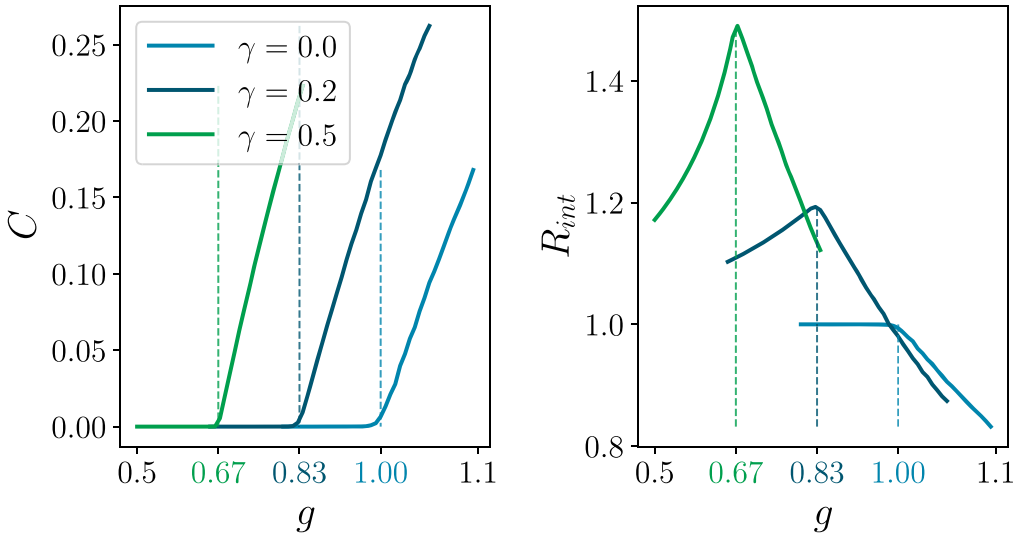

Figure 5. The behavior of correlation function C and response function R_{int} derived from the DMFT method [15]. We choose $\gamma = 0.0$, 0.2, 0.5, and g varies from 0.5 to 1.1 in the simulations. The vertical dashed lines denote the critical points for each γ value.

where $R(t, t') = \frac{\partial\langle\phi(x(t))\rangle}{\partial I(t')}$ is the two-time response function (I is an external perturbation), $\langle\omega(t)\omega(t')\rangle = g^2C(t, t')$, and $C(t, t') = \langle\phi(t)\phi(t')\rangle$ where the average is done in the dynamical mean-field sense.

In the steady-state, we have the property of time translation invariance $R(t, t') = R(t - t')$, and thus we can simplify the second term in the right-hand side of equation (14) as follows,

$$\int_0^\infty R(t, t')\phi(t')\mathrm{d}t' = \int_0^\infty R(u)\phi(t - u)\mathrm{d}u \to R_{\text{int}}\phi^*, \tag{15}$$

where $R_{\text{int}} = \int_0^\infty R(u)\mathrm{d}u$ is the time integral of $R(t)$, and ϕ^* denotes the steady-state value. Finally, we will obtain the following steady-state solution,

$$x^* = \omega^* + w\phi^*, \tag{16}$$

where $w \equiv g^2\gamma R_{\text{int}}$, and ω^* denotes a noise sampled from $\mathcal{N}(0, g^2C)$. For each noise sample, x^* and $\phi^* \equiv \phi(x^*)$ denote the optimal values satisfying equation (16). Therefore, correlation and response functions are self-consistently given by

$$\begin{aligned} C &= \langle(\phi^*)^2\rangle_\omega, \\ R_{\text{int}} &= \langle\phi'(\omega^* + w\phi^*)(1 + wR_{\text{int}})\rangle_\omega, \end{aligned} \tag{17}$$

where the average is done by drawing noise samples from $\mathcal{N}(0, g^2C)$.

A numerical iteration of equation (17) gives us the fixed-point of (C, R_{int}), which is plotted in figure 5. This corresponds exactly to the behavior of q (or Q), and even the behavior of the response order parameter derived using our quasi-potential method. We remark that our method takes an optimization angle, rather than the path integral angle. Therefore, it is reasonable to define and derive the order parameters in equilibrium ensemble theory (such as Boltzmann canonical ensemble here) and conclude the order-to-chaos transition is of the continuous type. Most interestingly, we can relate the replica overlap matrix to the observable neural activity, e.g., $q - Q$ measures the activity fluctuation, and $r - R$ measures the response of the system.

In fact, r and R display a peak at the transition point. This peak is an indicator of continuous dynamics transition and may be related to the divergence of the relaxation time-scale of the auto-correlation function [18, 19]. Most interestingly, the decay behavior of the response function on both sides of the transition looks very different, and on the chaotic side, the response decays more slowly. An earlier work revealed that the memory lifetime of the network under a linear decoder decreases more slowly on the chaotic side than on the

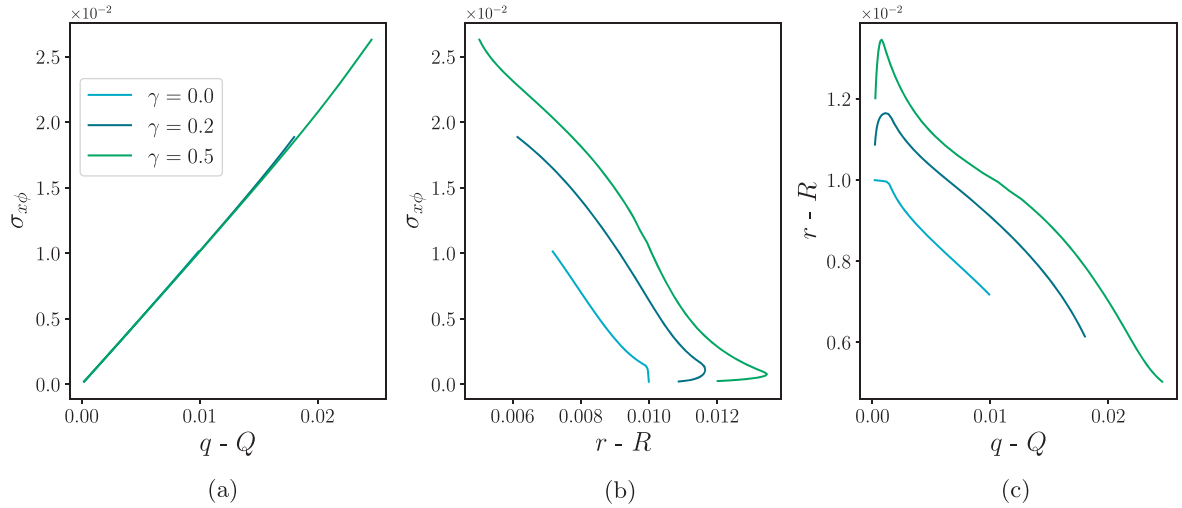

Figure 6. Fluctuation-response relationship for non-equilibrium steady states. The plotted order parameters follow the saddle-point equation derived in the main text [see equation (13*a*)]. (a) $\sigma_{x\phi}$ versus $q - Q$. (b) $\sigma_{x\phi}$ versus $r - R$. (c) $r - R$ versus $q - Q$.

non-chaotic side [30]. This shows that the response order parameters revealed in our current work have a deep connection to other performance metrics of random neural networks, and they may share the same mathematical foundation. We thus conclude that the dynamics at the edge of chaos are more responsive than at other locations in the phase diagram. The edge of chaos has thus great computational benefits, even a bit far away from the transition point to the chaotic regime [17, 31, 32]. Surprisingly, this is also consistent with the experimental data of cortical electrodynamics [33], which claimed the information richness occurs at the edge of chaos, and the conscious brain states are located near to the edge of chaos but not deep in the chaotic regime (see figure 2 in the recent work [33]). We conclude that this peak of the response function is a meaningful byproduct of our theory, and moreover an indicator of the edge of chaos.

A careful inspection of the saddle-point equation [see equation (13*a*)] leads to the following concise relationship between $q - Q$ (fluctuation) and $r - R$ (response):

$$r - R = \frac{\sqrt{\beta}\sigma_{x\phi}}{1 + \beta g^2 (q - Q)(1 + \gamma)}, \tag{18}$$

where $\sigma_{x\phi} = [\langle x\phi \rangle] - [\langle x \rangle \langle \phi \rangle]$, characterizing the current-activity correlation. Interestingly, the current-activity correlation behaves as an approximately linear function of the activity variance (figure 6). The relation between the response and fluctuation is quite non-linear, and at some location, there exhibits a peak that indicates the transition (figure 6). It is thus very interesting in future works to derive a non-equilibrium linear response theory.

4.3. Scaling behavior of order parameters in the zero temperature limit

In the zero temperature limit, we have to consider the following scaling behavior guaranteeing a finite value of free energy when $\beta \to \infty$ [see equation (11)].

$$\begin{aligned} q - Q &\to \frac{\chi}{\beta}, \quad \hat{Q} \to \beta^2 \hat{Q}, \\ 2\hat{q} - \hat{Q} &\to -2\beta\hat{\chi}, \quad r \to \sqrt{\beta}\tilde{r}, \\ R - r &\to \frac{\xi}{\sqrt{\beta}}, \quad \hat{R} \to \beta^{\frac{3}{2}}\kappa, \\ \hat{r} - \hat{R} &\to \beta\Gamma, \end{aligned} \tag{19}$$

where the new set of order parameters includes q, χ, $\hat{Q}$, $\hat{\chi}$, $\tilde{r}$, ξ, κ, and Γ, which are all of order one in magnitude. In the above scaling, one can derive a finite value of free energy density, from which the saddle point equation can also be derived (details are shown in appendix F).

The qualitative behavior of the new order parameter set is similar to that observed in the large β case (figure 7). The activity order parameter q becomes non-zero when the transition point is crossed to right-hand side, while χ has a sharp decrease when g goes below the transition point, implying the activity variance is strongly reduced when the dynamics have stable solutions. The behavior of $\tilde{r}$ is qualitatively the same as that of q in the shown regime, but we remark that they will develop a peak before reaching zero again with further increasing g. The order parameter $-\xi = \lim_{\beta\to\infty}\sqrt{\beta}(r - R)$ displays a peak at the transition point.

5. Concluding remarks

In this paper, we provide a relatively simple way to study non-equilibrium dynamics, by proposing a quasi-potential. Under this potential, we derive the stochastic dynamics to implement the optimization of searching for the low-speed region of the phase space. In the zero temperature limit, only the fixed points are considered, and thus our canonical ensemble theory yields a good description of the original non-potential dynamics in the long time limit. More precisely, the

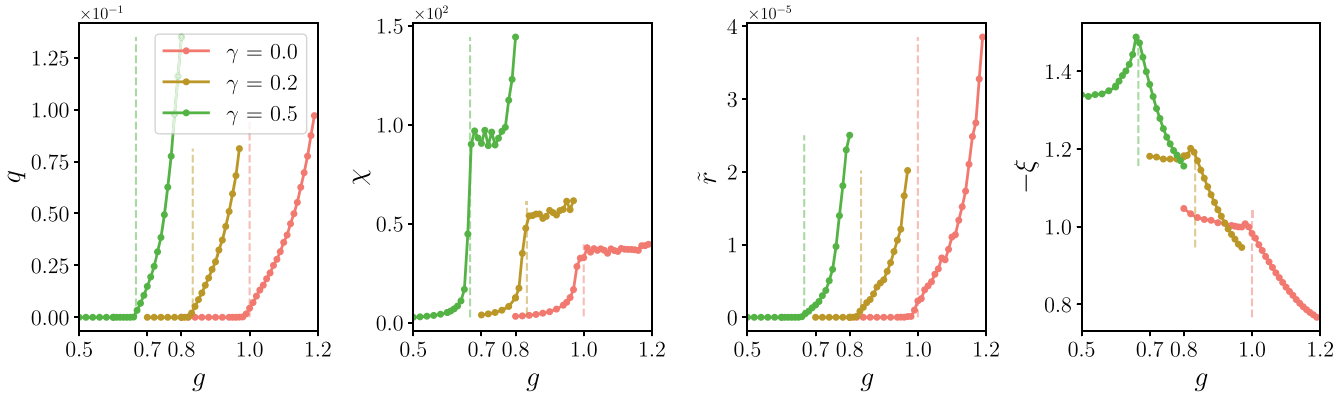

Figure 7. Order parameters as a function of increasing g for different values of γ. The results are obtained by solving the saddle-point equation in the zero temperature limit [see equation (F6) in appendix F].

low (zero)-speed landscape of the phase space can be analytically studied, which is relevant to our mechanistic understanding of recurrent neural dynamics, which can be adaptive, hierarchical, and noisy. This equilibrium transformation allows for using the Boltzmann measure to describe the low-speed landscape, as could be derived from the Fokker–Planck equation of the corresponding stochastic dynamics of the optimization [1]. Due to the quenched disorder present in our model, the replica method can be readily adapted as well.

The new finding is that two physically relevant order parameters are derived from our canonical ensemble calculation. One is activity-related, which exactly reproduces the order-to-chaos transition phenomenon, previously derived by using more complicated DMFT and random matrix theory [15, 18, 19, 25]. The other is response related, which shows a peak at the very transition point, demonstrating the computation benefits of the edge of chaos. The continuous nature of this chaos transition is confirmed by our theory as well. This transition is related to supersymmetry breaking in a field-theoretic representation of the dynamics in an action [34]. Remarkably, we relate the replica overlap matrix to measurable fluctuations and responses in non-equilibrium steady states of non-gradient neural dynamics.

Our work reveals a quasi-potential landscape behind non-gradient Langevin dynamics (or deterministic dynamics). This could inspire several important future directions. First, a complete picture of this landscape can be studied, as powerful equilibrium approaches can be leveraged, such as large deviation analysis [21], and landscape geometry analysis [35]. It would be very interesting to study the relationship between the distribution of equilibria and the onset of chaos in our framework, which we leave for a forthcoming paper [22]. Second, one can ask what is the quasi-potential of a coupled-dynamics system where the dynamics of constituent elements can be divided into several levels of different time scales. An enlarged state space is required to be introduced. Lastly, one could modify the quasi-potential dynamics to approximate the original non-gradient dynamics with increasing precision.

To sum up, our quasi-potential method provides a simple yet powerful tool to understand the non-equilibrium steady states, especially those without the equilibrium limit (non-gradient stochastic dynamics). Our method allows for order parameters to be defined in the canonical ensemble, determination of the transition type, and characterization of fluctuation-response relationship around the dynamics transition, thereby laying down a mathematical foundation of steady-state landscape of high-dimensional neural dynamics. We build a solid bridge from equilibrium ensemble theory to non-equilibrium non-gradient dynamics, and this optimization angle as a fresh route toward understanding non-equilibrium brain dynamics would prove fruitful in future studies.

Our approach does not rely on the special form of non-linear dynamics, as long as the speed of the dynamics can be identified, which already includes a broad range of neural dynamics models [17, 36]. In our framework, it would be very promising to treat the competition dynamics of many species in ecosystems [37–39], high-dimensional dynamics of random optimization [40, 41], high-dimensional gradient flow in machine learning [42], and adaptive, hierarchical and noisy neural dynamics in neuroscience [29, 36, 43–45]. The only limitation is that the transient behavior before the steady-state is reached and the steady-state dynamics could not be captured by our equilibrium approximation, and in that case, integro-differential equations or time-translation-invariant dynamics are required to solve with expensive computational cost [15]. Even if the dynamics contain slowly evolving couplings, we can also treat this as a quenched disorder that can be averaged resulting in a self-averaged free energy, from which transitions can be detected and analyzed. As some types of dynamics can not be accessed by DMFT, or the resultant integro-differential equations for correlation and response functions are hard to solve, our quasi-potential method is a promising alternative starting point to understanding the steady-state landscape of general non-equilibrium dynamics [46].

Acknowledgments

We would like to thank Zijian Jiang, Yang Zhao and Wenkang Du for the helpful discussions in earlier stages of this project. This research was supported by the National Natural Science Foundation of China under Grant No. 12122515 (HH), and Guangdong Provincial Key Laboratory of Magnetoelectric

Physics and Devices (Grant No. 2022B1212010008), and Guangdong Basic and Applied Basic Research Foundation (Grant No. 2023B1515040023).

Appendix A. Details of replica calculation

Starting from equation (8), we perform the Hubbard–Stratonovich (HS) transformation $e^{-\frac{1}{2}ab^2} = \int D\hat{x} e^{i\sqrt{a}b\hat{x}}$ for $a > 0$, to linearize the quadratic terms and have

$$Z^n = \int d\boldsymbol{x} D\hat{\boldsymbol{x}} \exp\left[i\sqrt{\beta}\sum_{ia} \hat{x}_i^a(-x_i^a + \sum_j J_{ij}\phi(x_j^a)\right) - \beta\eta\sum_a \|\boldsymbol{x}^a\|^2 \Bigg], \tag{A1}$$

where we define $D\hat{x} \equiv e^{-\frac{1}{2}\hat{x}^2} d\hat{x}/\sqrt{2\pi}$ as the Gaussian measure, and we call $\hat{x}$ as a response field [15]. We define i, j as the neuron index and a, b as the replica index. The neuron index runs from 1 to N, while the replica index runs from 1 to n. In addition, $d\boldsymbol{x} \equiv \prod_{i,a} dx_i^a$ and $D\hat{\boldsymbol{x}} \equiv \prod_{i,a} D\hat{x}_i^a$.

Before implementation of the quenched disorder average over the coupling statistics, it is convenient to separate the connection matrix into a symmetric part $\boldsymbol{J}^S$ and an anti-symmetric part $\boldsymbol{J}^A$ [12],

$$\boldsymbol{J} = \boldsymbol{J}^S + \rho\boldsymbol{J}^A, \tag{A2}$$

where by definition $J_{ij}^S = J_{ji}^S$ and $J_{ij}^A = -J_{ji}^A$, and their variances are defined as follows,

$$\langle (J_{ij}^S)^2 \rangle = \langle (J_{ij}^A)^2 \rangle = \frac{g^2}{N}\frac{1}{1+\rho^2}. \tag{A3}$$

Both parts are uncorrelated. It then follows that

$$\langle J_{ij}J_{ji} \rangle = \frac{g^2}{N}\frac{1-\rho^2}{1+\rho^2} = \frac{g^2}{N}\gamma, \tag{A4}$$

where the relationship between ρ and γ is given below,

$$\rho = \sqrt{\frac{1-\gamma}{1+\gamma}}. \tag{A5}$$

Defining $\langle\cdot\rangle$ as the quenched disorder average, we have

$$\langle Z^n \rangle = \int d\boldsymbol{x} D\hat{\boldsymbol{x}} \exp\left[-i\sqrt{\beta}\sum_{ia} x_i^a \hat{x}_i^a - \beta\eta\sum_a \|\boldsymbol{x}^a\|^2 \right] \times \left\langle \exp\left[i\sqrt{\beta}\sum_{ia} \hat{x}_i^a \sum_j J_{ij}\phi(x_j^a)\right]\right\rangle. \tag{A6}$$

The quenched disorder average is first completed as follows,

$$\begin{aligned}
&\left\langle \exp\left[i\sqrt{\beta}\sum_{ij} J_{ij}\sum_a \hat{x}_i^a\phi(x_j^a)\right]\right\rangle \\
&= \left\langle \exp\left[i\sqrt{\beta}\sum_{ij}(J_{ij}^S + \rho J_{ij}^A)\sum_a \hat{x}_i^a\phi(x_j^a)\right]\right\rangle \\
&= \exp\left[-\frac{1}{2}\beta\frac{g^2}{N}\frac{1}{1+\rho^2}\sum_{i<j}\sum_{ab}(\hat{x}_i^a\hat{x}_i^b\phi(x_j^a)\phi(x_j^b) + \hat{x}_i^a\hat{x}_j^b\phi(x_j^a)\phi(x_i^b) + \hat{x}_j^a\hat{x}_i^b\phi(x_i^a)\phi(x_j^b) + \hat{x}_j^a\hat{x}_j^b\phi(x_i^a)\phi(x_i^b))\right] \\
&\quad \times \exp\left[-\frac{1}{2}\beta\frac{g^2}{N}\frac{\rho^2}{1+\rho^2}\sum_{i<j}\sum_{ab}(\hat{x}_i^a\hat{x}_i^b\phi(x_j^a)\phi(x_j^b) - \hat{x}_i^a\hat{x}_j^b\phi(x_j^a)\phi(x_i^b) - \hat{x}_j^a\hat{x}_i^b\phi(x_i^a)\phi(x_j^b) + \hat{x}_j^a\hat{x}_j^b\phi(x_i^a)\phi(x_i^b))\right] \\
&= \exp\left[-\frac{1}{2}\beta\frac{g^2}{N}\sum_{i\neq j}\sum_{ab}\left(\frac{1}{1+\rho^2}(\hat{x}_i^a\hat{x}_i^b\phi(x_j^a)\phi(x_j^b) + \hat{x}_i^a\hat{x}_j^b\phi(x_j^a)\phi(x_i^b)) + \frac{\rho^2}{1+\rho^2}(\hat{x}_i^a\hat{x}_i^b\phi(x_j^a)\phi(x_j^b - \hat{x}_i^a\hat{x}_j^b\phi(x_j^a)\phi(x_i^b))\right)\right] \\
&= \exp\left[-\frac{1}{2}\beta\frac{g^2}{N}\sum_{ij}\sum_{ab}(\hat{x}_i^a\hat{x}_i^b\phi(x_j^a)\phi(x_j^b) + \gamma\hat{x}_i^a\hat{x}_j^b\phi(x_j^a)\phi(x_i^b))\right] \\
&= \exp\left[-\frac{1}{2}\beta g^2\sum_{ab} Q^{ab}\sum_i \hat{x}_i^a\hat{x}_i^b - \frac{1}{2}\beta g^2 N\gamma\sum_{ab} R^{ab}R^{ba}\right],
\end{aligned} \tag{A7}$$

where we are left with only γ (whose value can now take from -1 to 1, although $\gamma = -1$ is ill-defined for ρ), and we treat in our model the diagonal elements of the connection matrix negligible compared to the contribution from off-diagonal elements of the connection matrix, because of large N limit. To arrive at the last step, we have to introduce the following order parameters

$$Q^{ab} = \frac{1}{N}\sum_i \phi(x_i^a)\phi(x_i^b), \tag{A8}$$

$$R^{ab} = \frac{1}{N}\sum_i \hat{x}_i^a \phi(x_i^b). \tag{A9}$$

Hence, we need to insert the following identity in equation (A6),

$$\begin{aligned}1 &= \prod_{a\leqslant b}\int \mathrm{d}Q^{ab}\delta\left(Q^{ab} - \frac{1}{N}\sum_i \phi(x_i^a)\phi(x_i^b)\right)\prod_{ab}\int \mathrm{d}R^{ab}\delta\left(R^{ab} - \frac{1}{N}\sum_i \hat{x}_i^a\phi(x_i^b)\right)\\ &= \int \frac{\mathrm{d}\boldsymbol{Q}\mathrm{d}\hat{\boldsymbol{Q}}\mathrm{d}\boldsymbol{R}\mathrm{d}\hat{\boldsymbol{R}}}{(2\pi)^D}\exp\left[-\mathrm{i}\sum_{a\leqslant b}Q^{ab}\hat{Q}^{ab} - \mathrm{i}\sum_{ab}R^{ab}\hat{R}^{ab} + \mathrm{i}\frac{1}{N}\sum_i\sum_{a\leqslant b}\hat{Q}^{ab}\phi(x_i^a)\phi(x_i^b) + \mathrm{i}\frac{1}{N}\sum_{ab}\hat{R}^{ab}\sum_i \hat{x}_i^a\phi(x_i^b)\right]\\ &= \int \frac{\mathrm{d}\boldsymbol{Q}\mathrm{d}\hat{\boldsymbol{Q}}\mathrm{d}\boldsymbol{R}\mathrm{d}\hat{\boldsymbol{R}}}{[2\pi(\mathrm{i}/N)^2]^D}\exp\left[-N\sum_{a\leqslant b}Q^{ab}\hat{Q}^{ab} - N\sum_{ab}R^{ab}\hat{R}^{ab} + \sum_i\sum_{a\leqslant b}\hat{Q}^{ab}\phi(x_i^a)\phi(x_i^b) + \mathrm{i}\sum_{ab}\hat{R}^{ab}\sum_i\hat{x}_i^a\phi(x_i^b)\right],\end{aligned} \tag{A10}$$

where $D = n^2 + n(n+1)/2$, and we apply the Fourier transformation of Dirac delta function by introducing the conjugated order parameters $\hat{Q}^{ab}$, $\hat{R}^{ab}$, and rescale the order parameters as $\hat{Q}^{ab} \to -\mathrm{i}N\hat{Q}^{ab}$, $R^{ab} \to -\mathrm{i}R^{ab}$ and $\hat{R}^{ab} \to N\hat{R}^{ab}$ in the last step. $\mathrm{d}\boldsymbol{Q}\mathrm{d}\hat{\boldsymbol{Q}}\mathrm{d}\boldsymbol{R}\mathrm{d}\hat{\boldsymbol{R}}$ is a shorthand for $\prod_{a\leqslant b}\mathrm{d}Q^{ab}\mathrm{d}\hat{Q}_{ab}\prod_{ab}\mathrm{d}R^{ab}\mathrm{d}\hat{R}^{ab}$.

As a result, equation (A6) becomes

$$\begin{aligned}\langle Z^n\rangle \propto \int &\mathrm{d}\boldsymbol{x}\mathrm{D}\hat{\boldsymbol{x}}\mathrm{d}\boldsymbol{Q}\mathrm{d}\hat{\boldsymbol{Q}}\mathrm{d}\boldsymbol{R}\mathrm{d}\hat{\boldsymbol{R}}\\ &\times\exp\left[-\mathrm{i}\sqrt{\beta}\sum_{ia}x_i^a\hat{x}_i^a - \beta\eta\sum_{ia}(x_i^a)^2\right]\\ &\times\exp\left[-\frac{1}{2}g^2\beta\sum_{ab}Q^{ab}\left(\sum_i\hat{x}_i^a\hat{x}_i^b\right)\right.\\ &+\frac{1}{2}\beta g^2N\gamma\sum_{ab}R^{ab}R^{ba}\\ &-N\sum_{a\leqslant b}Q^{ab}\hat{Q}^{ab} - N\sum_{ab}R^{ab}\hat{R}^{ab}\\ &\left.+\sum_i\sum_{a\leqslant b}\hat{Q}^{ab}\phi(x_i^a)\phi(x_i^b) + \mathrm{i}\sum_{ab}\hat{R}^{ab}\sum_i\hat{x}_i^a\phi(x_i^b)\right],\end{aligned} \tag{A11}$$

where we have neglected the irrelevant constant due to the variable change in the integral, which does not affect the later saddle point approximation in the large N limit.

Equation (A11) can be rewritten into a compact form,

$$\begin{aligned}\langle Z^n\rangle \propto &\int \mathrm{d}\boldsymbol{Q}\mathrm{d}\hat{\boldsymbol{Q}}\mathrm{d}\boldsymbol{R}\mathrm{d}\hat{\boldsymbol{R}}\\ &\times\exp\left[N\left(-\sum_{a\leqslant b}Q^{ab}\hat{Q}^{ab} - \sum_{ab}R^{ab}\hat{R}^{ab} + \mathcal{G}\right)\right],\end{aligned} \tag{A12}$$

where the model-dependent action $\mathcal{G}$ reads,

$$\begin{aligned}\mathcal{G} = \ln\int &\mathrm{d}\boldsymbol{x}\mathrm{D}\hat{\boldsymbol{x}}\exp\left[-\mathrm{i}\sqrt{\beta}\sum_a x^a\hat{x}^a - \beta\eta\sum_a(x^a)^2\right.\\ &-\frac{1}{2}g^2\beta\sum_{ab}Q^{ab}\hat{x}^a\hat{x}^b + \sum_{a\leqslant b}\hat{Q}^{ab}\phi(x^a)\phi(x^b)\\ &\left.+\frac{1}{2}\beta g^2\gamma\sum_{ab}R^{ab}R^{ba} + \mathrm{i}\sum_{ab}\hat{R}^{ab}\hat{x}^a\phi(x^b)\right],\end{aligned} \tag{A13}$$

where $\mathrm{d}\boldsymbol{x}$ and $\mathrm{D}\hat{\boldsymbol{x}}$ have been reduced to $\prod_a\mathrm{d}x^a$ and $\prod_a\mathrm{D}\hat{x}^a$, respectively.

To proceed, we have to assume the replica symmetric (RS) ansätz as the first level of approximation, i.e., the replica overlap matrix is permutation invariant, as is intuitively expected from the replicated partition function. This first level of approximation can be refined by going to multiple steps of replica symmetry breaking [20], provided that the theoretical prediction is in significant disagreement with numerical simulations. The RS ansätz reads $Q^{ab} = q\delta_{ab} + Q(1-\delta_{ab})$, $\hat{Q}^{ab} = \hat{q}\delta_{ab} + \hat{Q}(1-\delta_{ab})$, $R^{ab} = r\delta_{ab} + R(1-\delta_{ab})$ and $\hat{R}^{ab} = \hat{r}\delta_{ab} + \hat{R}(1-\delta_{ab})$. We have then the following result of the averaged replicated partition function,

$$\begin{aligned}\langle Z^n\rangle \propto &\int \mathrm{d}Q\mathrm{d}\hat{Q}\mathrm{d}q\mathrm{d}\hat{q}\mathrm{d}R\mathrm{d}\hat{R}\mathrm{d}r\mathrm{d}\hat{r}\\ &\times\exp\left[N\left(-\frac{n(n-1)}{2}Q\hat{Q} - nq\hat{q} - n(n-1)R\hat{R} - nr\hat{r} + \mathcal{G}\right)\right],\end{aligned} \tag{A14}$$

where

$$\begin{aligned}\mathcal{G} = \ln\int &\mathrm{d}\boldsymbol{x}\mathrm{D}\hat{\boldsymbol{x}}\exp\left[-\mathrm{i}\sqrt{\beta}\sum_a x^a\hat{x}^a - \beta\eta\sum_a(x^a)^2\right.\\ &+\frac{1}{2}\beta g^2\gamma(n(n-1)R^2 + nr^2)\\ &-\frac{1}{2}g^2\beta\left(Q\left(\sum_a\hat{x}^a\right)^2 + (q-Q)\sum_a(\hat{x}^a)^2\right)\\ &+\frac{1}{2}\left(\hat{Q}\left(\sum_a\phi(x^a)\right)^2 + (2\hat{q}-\hat{Q})\sum_a\phi^2(x^a)\right)\\ &\left.+\mathrm{i}\left(\hat{R}\sum_{ab}\hat{x}^a\phi(x^b) + (\hat{r}-\hat{R})\sum_a\hat{x}^a\phi(x^a)\right)\right].\end{aligned} \tag{A15}$$

To linearize the quadratic terms in equation (A15), we use the HS transformation once again as follows,

$$\begin{aligned}&\mathrm{e}^{\frac{1}{2}\hat{Q}\left(\sum_a\phi(x^a)\right)^2 - \frac{1}{2}g^2\beta Q\left(\sum_a\hat{x}^a\right)^2 + \mathrm{i}\hat{R}\sum_{ab}\hat{x}^a\phi(x^b)}\\ &= \int \mathrm{D}u\mathrm{D}v\mathrm{e}^{\kappa_1 u\sum_a\phi(x^a) + v\left(\kappa_2\sum_a\phi(x^a) + \mathrm{i}\kappa_3\sum_a\hat{x}^a\right)},\end{aligned} \tag{A16}$$

where the auxiliary coefficients $\kappa_1 = \sqrt{\frac{g^2\beta Q\hat{Q} - \hat{R}^2}{g^2\beta Q}}$, $\kappa_2 = \frac{\hat{R}}{g\sqrt{\beta Q}}$ and $\kappa_3 = g\sqrt{\beta Q}$.

Therefore, we have

$$\mathcal{G} = \ln \int \mathrm{D}u\mathrm{D}v \exp\left[\frac{1}{2}\beta g^2\gamma(n(n-1)R^2 + nr^2)\right] I^n, \tag{A17}$$

where

$$\begin{aligned} I \equiv \int \mathrm{d}x\mathrm{D}\hat{x} \exp\Big[& -\beta\eta x^2 - \mathrm{i}\sqrt{\beta}x\hat{x} - \frac{1}{2}g^2\beta(q-Q)\hat{x}^2 \\ & + \frac{1}{2}(2\hat{q} - \hat{Q})\phi^2(x) + \mathrm{i}(\hat{r} - \hat{R})\hat{x}\phi(x) \\ & + \mathrm{i}\kappa_3 v\hat{x} + (\kappa_1 u + \kappa_2 v)\phi(x)\Big]. \end{aligned} \tag{A18}$$

We next complete the integral over $\mathrm{D}\hat{x}$ as

$$\begin{aligned} & \int \mathrm{D}\hat{x} \exp\left[-\frac{1}{2}g^2\beta(q-Q)\hat{x}^2 + \mathrm{i}(\kappa_3 v + (\hat{r} - \hat{R}) - \sqrt{\beta}x)\hat{x}\right] \\ & = \frac{1}{\sigma}\exp\left[-\frac{1}{2}\frac{1}{\sigma^2}(\kappa_3 v + (\hat{r} - \hat{R})\phi(x) - \sqrt{\beta}x)^2\right], \end{aligned} \tag{A19}$$

where $\sigma \equiv \sqrt{1 + g^2\beta(q-Q)}$. Thus we obtain,

$$\begin{aligned} \mathcal{G} = & -n\ln\sigma + \frac{1}{2}\beta g^2\gamma(n(n-1)R^2 + nr^2) \\ & + \ln \int \mathrm{D}u\mathrm{D}v I^n, \end{aligned} \tag{A20}$$

where $I \equiv \int \mathrm{d}x \mathrm{e}^{\mathcal{H}(x)}$, and

$$\begin{aligned} \mathcal{H}(x) \equiv & -\beta\eta x^2 + \frac{1}{2}(2\hat{q} - \hat{Q})\phi^2(x) + (\kappa_1 u + \kappa_2 v)\phi(x) \\ & - \frac{1}{2}\frac{1}{\sigma^2}(\kappa_3 v + (\hat{r} - \hat{R})\phi(x) - \sqrt{\beta}x)^2, \end{aligned} \tag{A21}$$

where $\mathcal{H}(x)$ could be considered as the single-variable effective Hamiltonian of our model, which emerges after the quenched disorder is averaged out.

Finally, we perform the saddle point approximation because of large N, such that $\langle Z^n\rangle \propto \int \mathrm{d}\mathcal{O}\mathrm{d}\hat{\mathcal{O}}\mathrm{e}^{N\Phi(\mathcal{O},\hat{\mathcal{O}})} \approx \mathrm{e}^{N\Phi(\mathcal{O}^*,\hat{\mathcal{O}}^*)}$, where $\mathcal{O} = \{Q, q, R, r\}$ and the superscript $*$ refers to the saddle point value. The free energy can be obtained from the replica trick $-\beta f = \lim_{n\to 0}\frac{\ln\langle Z^n\rangle}{nN}$, and therefore,

$$\begin{aligned} -\beta f = & \frac{1}{2}Q\hat{Q} - q\hat{q} + R\hat{R} - r\hat{r} \\ & - \ln\sigma + \frac{1}{2}\beta g^2\gamma(r^2 - R^2) + \int \mathrm{D}u\mathrm{D}v \ln I. \end{aligned} \tag{A22}$$

Appendix B. Saddle-point equations

In this section, we omit the superscript $*$ on the order parameters and derive the self-consistent equations these order parameters must obey, which are the so-called saddle point equations (SDEs). The principle is to optimize the action with respect to the order parameters and their conjugated counterparts. The result is given below,

$$q = [\langle\phi^2\rangle], \tag{B1a}$$

$$Q = [\langle\phi\rangle^2], \tag{B1b}$$

$$\begin{aligned} r = & -\frac{1}{\sigma^2}f(1,0,-1)[\langle\phi^2\rangle] + \frac{1}{\sigma^2}f(0,1,-1)[\langle\phi\rangle^2] \\ & + \frac{\sqrt{\beta}}{\sigma^2}(1 - gkQ)[\langle x\phi\rangle] + \frac{\sqrt{\beta}}{\sigma^2}gkQ[\langle x\rangle\langle\phi\rangle], \end{aligned} \tag{B1c}$$

$$\begin{aligned} R = & -\frac{1}{\sigma^2}f(0,1,-1)[\langle\phi^2\rangle] - \frac{1}{\sigma^2}f(1,-2,1)[\langle\phi\rangle^2] \\ & - \frac{\sqrt{\beta}}{\sigma^2}gkQ[\langle x\phi\rangle] + \frac{\sqrt{\beta}}{\sigma^2}(1 + gkQ)[\langle x\rangle\langle\phi\rangle], \end{aligned} \tag{B1d}$$

$$\begin{aligned} \hat{q} = & -\frac{gk}{2} + \frac{g^2k^2Q}{2} + \frac{gk}{2\sigma^2}(\hat{r} - \hat{R})f(1,1,-2)[\langle\phi^2\rangle] \\ & + \frac{gk}{\sigma^2}(\hat{r} - \hat{R})f(0,-1,1)[\langle\phi\rangle^2] + \frac{k^2}{2}(1 - 2gkQ)[\langle x^2\rangle] \\ & + \frac{gk\sqrt{\beta}}{\sigma^2}f(0,1,-2)[\langle x\rangle\langle\phi\rangle] \\ & + \frac{gk\sqrt{\beta}}{\sigma^2}f(-1,0,2)[\langle x\phi\rangle] + gk^3Q[\langle x\rangle^2], \end{aligned} \tag{B1e}$$

$$\begin{aligned} \hat{Q} = & g^2k^2Q + \frac{2gk}{\sigma^2}(\hat{r} - \hat{R})f(0,1,-1)[\langle\phi^2\rangle] \\ & + \frac{gk}{\sigma^2}(\hat{r} - \hat{R})f(1,-3,2)[\langle\phi\rangle^2] - 2gk^3Q[\langle x^2\rangle] \\ & - \frac{2gk\sqrt{\beta}}{\sigma^2}f(1,-2,2)[\langle x\rangle\langle\phi\rangle] \\ & + \frac{2gk\sqrt{\beta}}{\sigma^2}f(0,-1,2)[\langle x\phi\rangle] \\ & + k^2(1 + 2gkQ)[\langle x\rangle^2], \end{aligned} \tag{B1f}$$

$$\hat{r} = \beta g^2\gamma r, \tag{B1g}$$

$$\hat{R} = \beta g^2\gamma R, \tag{B1h}$$

where $k \equiv \frac{g\beta}{\sigma^2}$ and $f(a,b,c) \equiv a\hat{r} + b\hat{R} + cgkQ(\hat{r} - \hat{R})$, and we define disorder and thermal averages, respectively as follows,

$$[\bullet] \equiv \int \mathrm{D}u\mathrm{D}v\bullet, \tag{B2}$$

and

$$\langle\bullet\rangle \equiv \frac{\int \mathrm{d}x\mathrm{e}^{\mathcal{H}(x)}\bullet}{\int \mathrm{d}x\mathrm{e}^{\mathcal{H}(x)}}. \tag{B3}$$

Note that we have used the integral identity $\int \mathrm{D}x x f(x) = \int \mathrm{D}x f'(x)$ where $f(x)$ is a bounded and differentiable function to write the above SDEs in the compact form.

We are also interested in the average energy, whose ground state value is related to the speed level of the original dynamics. The thermodynamic relation $\langle E\rangle = \frac{\partial[\beta f]}{\partial\beta}$ is used to

derive the following formula,

$$\begin{aligned}U = &\frac{g^2}{2\sigma^2}(q - gkQ(q - Q)) - \frac{1}{2}g^2(r^2 - R^2)\gamma \\ &+ \frac{1}{2\sigma^2}\left(1 + 2\eta\sigma^2 - gk(q - Q) - \frac{2gkQ}{\sigma^2}\right)[\langle x^2\rangle] + \frac{gkQ}{\sigma^4}[\langle x\rangle^2] \\ &- \frac{1}{2\sqrt{\beta}\sigma^2}\left(f(1, -1, 0) + \frac{1}{\sigma^2}f(0, 1, -3)\right. \\ &\left.+ gk(q - Q)f(-2, 1, 1)\right)[\langle x\phi\rangle] \\ &- \frac{1}{2\sqrt{\beta}\sigma^2}\left(\frac{1}{\sigma^2}f(0, -1, 3) + gk(q - Q)f(0, 1, -1)\right)[\langle x\rangle\langle\phi\rangle] \\ &- \frac{1}{2g^2Q\beta^2}(\hat{R}^2 + g^2k^2Q(q - Q)(\hat{r} - \hat{R})^2 \\ &+ f(0, -1, 1)(f(0, 1, 1) - 2g^2k^2Q(q - Q)(\hat{r} - \hat{R})))[\langle\phi^2\rangle] \\ &+ \frac{1}{2g^2Q\beta^2}(\hat{R}^2 + f(0, -1, 1)(f(0, 1, 1) - 2g^2k^2Q(q - Q)(\hat{r} - \hat{R})))[\langle\phi\rangle^2]. \end{aligned} \tag{B4}$$

To characterize the average activity level in the neural population, we also derive the ℓ_2 norm of the activity as follows

$$\frac{1}{N}\langle\|\boldsymbol{x}\|^2\rangle = -\frac{1}{\beta}\frac{\partial(-\beta f)}{\partial\eta} = [\langle x^2\rangle]. \tag{B5}$$

Appendix C. Independent case $\gamma = 0$

In the case $\gamma = 0$, all the connection weights are independent and uncorrelated. Hence, the expression of the SDEs gets greatly simplified as

$$q = [\langle\phi^2\rangle], \tag{C1}$$

$$Q = [\langle\phi\rangle^2], \tag{C2}$$

$$r = \frac{\sqrt{\beta}}{\sigma^2}(1 - gkQ)[\langle x\phi\rangle] + \frac{\sqrt{\beta}}{\sigma^2}gkQ[\langle x\rangle\langle\phi\rangle], \tag{C3}$$

$$R = -\frac{\sqrt{\beta}}{\sigma^2}gkQ[\langle x\phi\rangle] + \frac{\sqrt{\beta}}{\sigma^2}(1 + gkQ)[\langle x\rangle\langle\phi\rangle], \tag{C4}$$

$$\begin{aligned}\hat{q} = &-\frac{gk}{2} + \frac{g^2k^2Q}{2} + gk^3Q[\langle x\rangle^2] \\ &+ \frac{k^2}{2}(1 - 2gkQ)[\langle x^2\rangle], \end{aligned} \tag{C5}$$

$$\hat{Q} = g^2k^2Q - 2gk^3Q[\langle x^2\rangle] + k^2(1 + 2gkQ)[\langle x\rangle^2], \tag{C6}$$

$$\hat{r} = 0, \tag{C7}$$

$$\hat{R} = 0. \tag{C8}$$

Because $\hat{r} = \hat{R} = 0$, we have $f(a, b, c) = 0$. The free energy is thus reduced to

$$-\beta f = \frac{1}{2}Q\hat{Q} - q\hat{q} - \ln\sigma - \frac{g^2\beta}{2\sigma^2}Q + \int \mathrm{D}u\mathrm{D}v\ln I, \tag{C9}$$

where

$$\begin{aligned}I \equiv &\int \mathrm{d}x\exp\left[-\frac{1}{2}\beta\left(\frac{1}{\sigma^2} + 2\eta\right)x^2\right] \\ &\times\exp\left[+\frac{1}{2}(2\hat{q} - \hat{Q})\phi^2(x) + \sqrt{\hat{Q}}u\phi(x) + \frac{g\beta}{\sigma^2}\sqrt{Q}vx\right]. \end{aligned} \tag{C10}$$

This free energy does not depend on the order parameters r and R anymore. The average energy also yields the following simple form

$$\begin{aligned}U = &\frac{g^2}{2\sigma^2}(q - gkQ(q - Q)) + \frac{gkQ}{\sigma^4}[\langle x\rangle^2] \\ &+ \frac{1}{2\sigma^2}\left(1 + 2\eta\sigma^2 - gk(q - Q) - \frac{2gkQ}{\sigma^2}\right)[\langle x^2\rangle]. \end{aligned} \tag{C11}$$

Appendix D. Additional experiment results

To illustrate the phase transition in more detail, we choose some curves ($\gamma = 0.2,\ 0.5,\ 0.8$) and plot them in the 2-dimension plane (figure 8). q and Q are close to zero before the critical point while growing rapidly after the critical point. r and R achieve their maximum values at the critical point.

Next, we explore the effect of β in optimizing the energy and the order parameters. When g is close to the critical point, we show how q varies with increasing β (see figure 9 for $g = 0.98$ and $g = 1.0$). We observe that q converges more rapidly towards zero when g is below the critical value. When g is approaching the critical value from below, the decaying towards zero becomes much slower, signaling a continuous phase transition. This is also supported by the iteration dynamics of the SDEs (figure 10).

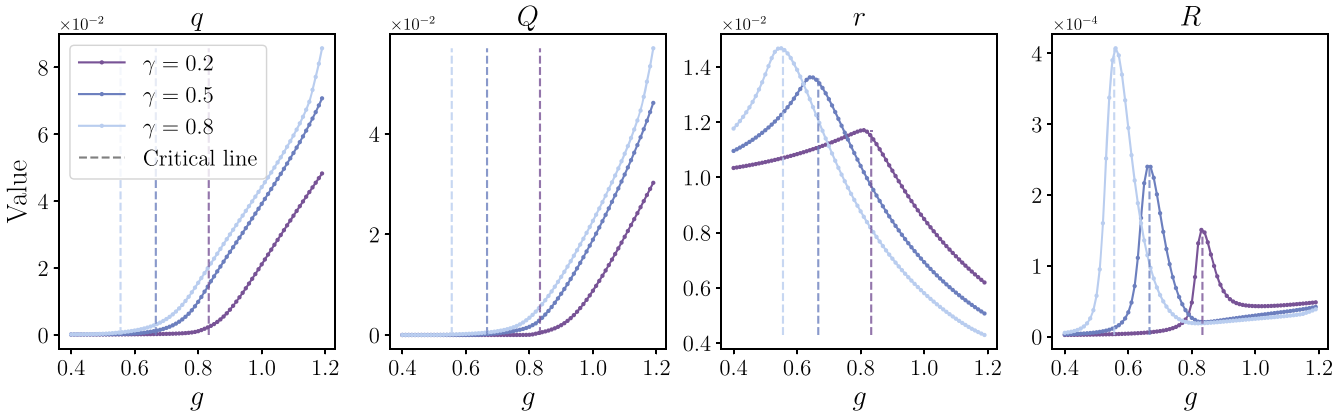

Figure 8. Phase transitions as g increases. We plot three different curves ($\gamma = 0.2, 0.5, 0.8$), and different colors represent different γ values. The dashed line represents the DMFT result of critical points in [19].

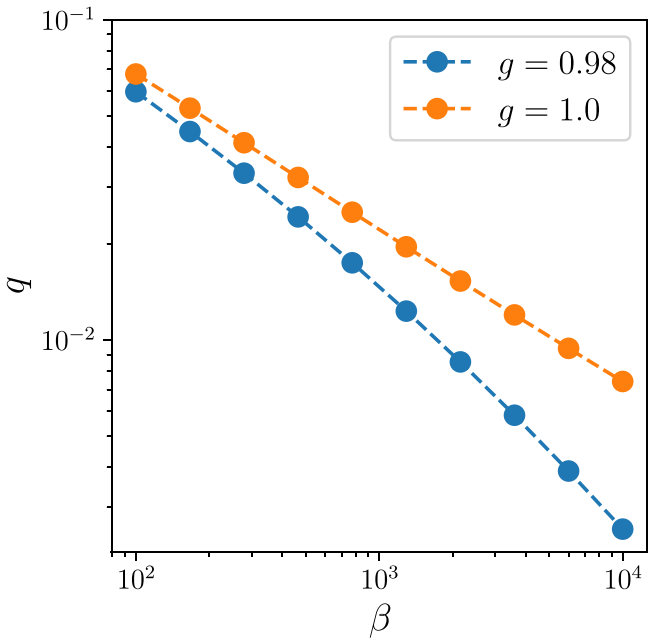

Figure 9. The relationship between q and β when $g = 0.98$ and $g = 1.0$ ($\gamma = 0$). The blue dots denote $g = 0.98$ case, and the orange dots denote $g = 1.0$. Both axes are in the log scale.

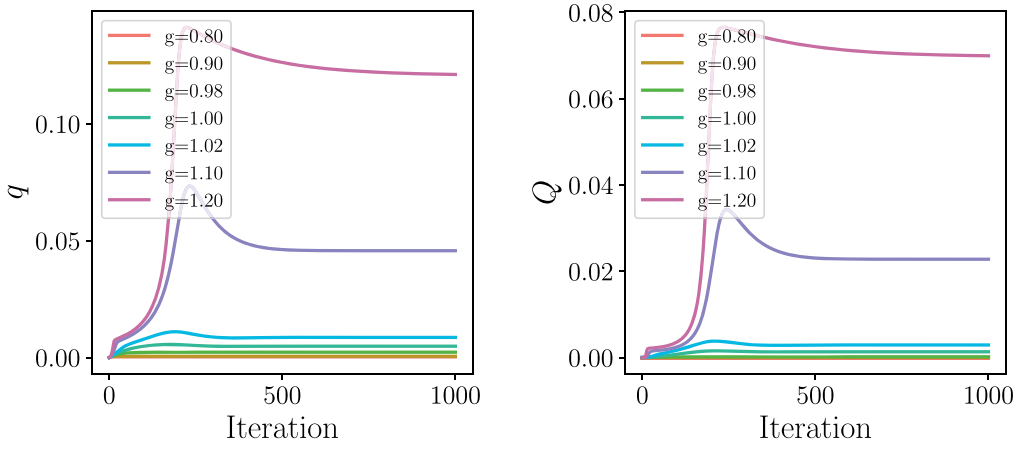

Figure 10. The iterative dynamics of q and Q for different values of g. All initial values are set to 10^{-5} for q and 10^{-6} for Q. The data points are obtained by using the Algo. 1.

Appendix E. Numerical details of solving SDEs

The most expensive cost of solving SDEs comes from the double average terms, such as $[\langle x\rangle^2]$ and $[\langle\phi\rangle^2]$, where the outer integral is related to Duv and the inner integral is related dx. These integrals are calculated by the Monte Carlo (MC) method with the important sampling technique [17]. Therefore, we need to generate $100k \times 6k$ Gaussian samples, where $100k$ MC samples of u and v are used for the outer integral and $6k$ MC samples of x for each pair of u and v, is used for the inner integral. k here represents 1 000. Note that the inner integral in equation (B3) does not contain the Gaussian measure at first sight. However, we can take a reorganization of terms in the effective Hamiltonian. We then get the following equivalent transformation,

$$\langle\bullet\rangle \equiv \frac{\int \mathrm{D}x \mathrm{e}^{\tilde{\mathcal{H}}(x)}\bullet}{\int \mathrm{D}x \mathrm{e}^{\tilde{\mathcal{H}}(x)}}, \tag{E1}$$

where the Gaussian measure is introduced as $\mathrm{D}x \equiv \mathrm{e}^{-\frac{1}{2}\beta\left(\frac{1}{\sigma^2}+2\eta\right)x^2}\mathrm{d}x\Big/\sqrt{2\pi}$ and

$$\begin{aligned}\tilde{\mathcal{H}}(x) \equiv& \frac{1}{2}(2\hat{q} - \hat{Q})\phi^2(x) + (\kappa_1 u + \kappa_2 v)\phi(x)\\ &- \frac{1}{2}\frac{1}{\sigma^2}(\kappa_3 v + (\hat{r} - \hat{R})\phi(x))^2\\ &+ \frac{\sqrt{\beta}}{\sigma^2}(\kappa_3 v + (\hat{r} - \hat{R})\phi(x))x.\end{aligned} \tag{E2}$$

After an initialization of the order parameters, we complete the double average using the above MC sampling. Then we update the order parameters as follows,

$$\mathcal{O}_{t\,+\,1} = \alpha\mathcal{O}_t + (1 - \alpha)f(\mathcal{O}_t), \tag{E3}$$

where α is the damping parameter (a value close to one) to speed up the convergence. $f(\mathcal{O})$ represents the right-hand-side of equations (B1a). We stop the iteration when $|\mathcal{O}_{t+1} - \mathcal{O}_t| < 10^{-3}$ for all order parameters. We give a pseudocode in Algorithm. 1 for solving the SDEs.

Algorithm 1. Solver of SDEs

Input: The initial values of order parameters $\mathcal{O}_0$, $\mathcal{O} \in \{q, Q, r, R\}$
Output: The convergent values of order parameters $\mathcal{O}^*$
1: Generate $100k \times 6k$ Gaussian Monte Carlo samples
2: **repeat**
3: Calculate all the double averages by the MC method in equation (B1a)
4: Let $\mathcal{O}_{t+1} \leftarrow \alpha\mathcal{O}_t + (1 - \alpha)f(\mathcal{O}_t)$, where $f(\mathcal{O})$ are the functions specified by the SDEs [see equation (B1a)]
5: **until** $|\mathcal{O}_{t+1} - \mathcal{O}_t| < 10^{-3}$
6: Let $\mathcal{O}^* \leftarrow \mathcal{O}_t$

Appendix F. Zero temperature analysis

Setting $\beta \to \infty$, we focus on the ground state of the quasi-potential, which describes the non-equilibrium steady states of equation (1). In the zero temperature limit, we have to consider the following scaling behavior guaranteeing a finite value of free energy when $\beta \to \infty$ [see equation (11)].

$$\begin{aligned} q - Q &\to \frac{\chi}{\beta},\\ \hat{Q} &\to \beta^2\hat{Q},\\ 2\hat{q} - \hat{Q} &\to -2\beta\hat{\chi},\\ r &\to \sqrt{\beta}\tilde{r},\\ R - r &\to \frac{\xi}{\sqrt{\beta}},\\ \hat{R} &\to \beta^{\frac{3}{2}}\kappa,\\ \hat{r} - \hat{R} &\to \beta\Gamma,\end{aligned} \tag{F1}$$

where the new set of order parameters includes q, χ, $\hat{Q}$, $\hat{\chi}$, $\tilde{r}$, ξ, κ, and Γ, which are all of order one in magnitude.

Inserting the above scaling expressions into the free energy [see equation (11)], which then reads

$$\begin{aligned}-\beta f =& -\frac{1}{2}\beta(-2q\hat{\chi} + \hat{Q}\chi) - \beta(\tilde{r}\Gamma - \kappa\xi)\\ &- g^2\beta\gamma\xi\tilde{r} + \int (\mathrm{D}u\mathrm{D}v)\ln\int \mathrm{d}x \mathrm{e}^{\beta\mathcal{H}_0(x)},\end{aligned} \tag{F2}$$

where

$$\begin{aligned}\mathcal{H}_0(x) =& -\eta x^2 - \hat{\chi}\phi^2(x)\\ &+ \left(\sqrt{\hat{Q} - \frac{\kappa^2}{g^2 q}}u + \frac{\kappa}{g\sqrt{q}}v\right)\phi(x)\\ &- \frac{1}{2}\frac{1}{\sigma^2}(g\sqrt{q}v + \Gamma\phi(x) - x)^2.\end{aligned} \tag{F3}$$

In the zero temperature limit $\beta \to \infty$, we apply the Laplace method to estimate the following integral

$$\int \mathrm{D}u\mathrm{D}v \ln\int \mathrm{d}x \mathrm{e}^{\beta\mathcal{H}_0(x)} = \beta\int \mathrm{D}u\mathrm{D}v\mathcal{H}_0(x^*), \tag{F4}$$

where $x^* = argmax_x\, \mathcal{H}_0(x)$. Note that all terms in $-\beta f$ are proportional to β, and hence the free energy density in the zero temperature limit is given by

$$\begin{aligned}-f =& -\frac{1}{2}(-2q\hat{\chi} + \hat{Q}\chi) - (\tilde{r}\Gamma - \kappa\xi) - g^2\gamma\xi\tilde{r}\\ &+ \int \mathrm{D}u\mathrm{D}v\mathcal{H}_0(x^*).\end{aligned} \tag{F5}$$

The new set of order parameters obeys the self-consistent equation making the zero temperature free energy stationary. We thus get the following saddle-point equation:

$$
\begin{aligned}
q &= [\phi^2(x^*)],\\
\chi &= \frac{gq}{\sqrt{g^2q^2\hat{Q} - q\kappa^2}}[u\phi(x^*)],\\
\hat{Q} &= \frac{g^2}{\sigma^4}(g^2q + \Gamma^2[\phi^2(x^*)] + [(x^*)^2]\\
&\quad - 2\Gamma[x^*\phi(x^*)] + 2g\sqrt{q}\Gamma[v\phi(x^*)] - 2g\sqrt{q}[vx^*]),\\
\hat{\chi} &= \frac{g^2}{2\sigma^2} - \frac{\kappa^2}{2gq\sqrt{\hat{Q}g^2q^2 - q\kappa^2}}[u\phi(x^*)]\\
&\quad + \frac{1}{2\sqrt{q}}\left(\frac{g\Gamma}{\sigma^2} + \frac{\kappa}{gq}\right)[v\phi(x^*)] - \frac{1}{\sigma^2}\frac{g}{2\sqrt{q}}[vx^*],\\
\tilde{r} &= -\frac{g\sqrt{q}}{\sigma^2}[v\phi(x^*)] - \frac{\Gamma}{\sigma^2}[\phi^2(x^*)] + \frac{1}{\sigma^2}[x^*\phi(x^*)],\\
\xi &= \frac{\kappa}{g\sqrt{g^2q^2\hat{Q} - q\kappa^2}}[u\phi(x^*)] - \frac{1}{g\sqrt{q}}[v\phi(x^*)],\\
\kappa &= g^2\gamma\tilde{r},\\
\Gamma &= -g^2\gamma\xi,
\end{aligned}
\tag{F6}
$$

where $[\bullet] \equiv \int \mathrm{D}u\mathrm{D}v$ as before, and to estimate this average, we first generate $M = 5\times10^6$ Monte Carlo samples $\{(u_i, v_i)\}_{i=1}^{M}$, for each of them we find the global maximum of $\mathcal{H}_0(x)$, i.e., x^*. All these values corresponding to the maxima [for each pair of (u, v)] are further used to complete the calculation of the order parameter for one round. The codes of this paper are released on the GitHub link [47].

References

[1] Risken H 1996 *The Fokker-Planck Equation: Methods of Solution and Applications* (Berlin: Springer)

[2] Strogatz S H 2015 *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Boca Raton, FL: CRC Press)

[3] Rabinovich M I, Varona P, Selverston A I and Abarbanel H D I 2006 Dynamical principles in neuroscience *Rev. Mod. Phys.* **78** 1213–65

[4] Kurchan J 1998 Fluctuation theorem for stochastic dynamics *J. Phys. A: Math. Gen.* **31** 3719

[5] Takahiro O and Ohta T 2007 Fluctuation theorems for non-linear generalized langevin systems *J. Stat. Mech: Theory Exp.* **2007** P10010

[6] Seifert U 2012 Stochastic thermodynamics, fluctuation theorems and molecular machines *Rep. Prog. Phys.* **75** 126001

[7] Martin P C, Siggia E D and Rose H A 1973 Statistical dynamics of classical systems *Phys. Rev.* A **8** 423–37

[8] Janssen H-K 1976 On a lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties *Zeitschrift für Physik B: Condensed Matter* **23** 377–80

[9] De Dominicis C 1978 Dynamics as a substitute for replicas in systems with quenched random impurities *Phys. Rev.* B **18** 4913–9

[10] Sommers H-J 1987 Path-integral approach to ising spin-glass dynamics *Phys. Rev. Lett.* **58** 1268–71

[11] Hatano T and Sasa S-i 2001 Steady-state thermodynamics of langevin systems *Phys. Rev. Lett.* **86** 3463–6

[12] Crisanti A and Sompolinsky H 1987 Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model *Phys. Rev.* A **36** 4922–39

[13] Chow C C and Buice M A 2015 Path integral methods for stochastic differential equations *J. Math. Neuro.* **5** 8

[14] Hertz J A, Roudi Y and Sollich P 2017 Path integral methods for the dynamics of stochastic and disordered systems *J. Phys. A: Math. Theor.* **50** 033001

[15] Zou W and Huang H 2024 Introduction to dynamical mean-field theory of randomly connected neural networks with bidirectionally correlated couplings *SciPost Phys. Lect. Notes* **79**

[16] Sussillo D and Barak O 2013 Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks *Neural Comput.* **25** 626–49

[17] Huang H 2022 *Statistical Mechanics of Neural Networks* (Springer)

[18] Sompolinsky H, Crisanti A and Sommers H J 1988 Chaos in random neural networks *Phys. Rev. Lett.* **61** 259–62

[19] Marti D, Brunel N and Ostojic S 2018 Correlations between synapses in pairs of neurons slow down dynamics in randomly connected neural networks *Phys. Rev.* E **97** 062314

[20] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (World Scientific)

[21] Touchette H 2009 The large deviation approach to statistical mechanics *Phys. Rep.* **478** 1–69

[22] Wang S and Huang H 2024 *In preparation*

[23] Amari S-I 1972 Characteristics of random nets of analog neuron-like elements *IEEE Transactions on Systems, Man, and Cybernetics, SMC* **2** 643–57

[24] Wainrib G and Touboul J 2013 Topological and dynamical complexity of random neural networks *Phys. Rev. Lett.* **110** 118101

[25] Sommers H J, Crisanti A, Sompolinsky H and Stein Y 1988 Spectrum of large random asymmetric matrices *Phys. Rev. Lett.* **60** 1895–8

[26] Ao P 2004 Potential in stochastic differential equations: novel construction *J. Phys. A: Math. Gen.* **37** L25

[27] Yan H, Zhao L, Hu L, Wang X, Wang E and Wang J 2013 Nonequilibrium landscape theory of neural networks *PNAS* **110** E4185–94

[28] Coolen A C C 2001 Statistical mechanics of recurrent neural networks i—statics *Handbook of Biological Physics* volume 4 (Elsevier) 553–618

[29] Helias M and Dahmen D 2020 *Statistical field theory for neural networks* (Springer)

[30] Toyoizumi T and Abbott L F 2011 Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime *Phys. Rev.* E **84** 051908

[31] Langton C G 1990 Computation at the edge of chaos: Phase transitions and emergent computation *Physica* D **42** 12–37

[32] Bertschinger N and Natschläger T 2004 Real-time computation at the edge of chaos in recurrent neural networks *Neural Comput.* **16** 1413–36

[33] Daniel Toker I *et al* 2022 Consciousness is supported by near-critical slow cortical electrodynamics *Proc. Natl Acad. Sci.* **119** e2024455119

[34] Parisi G and Sourlas N 1982 Supersymmetric field theories and stochastic differential equations *Nucl. Phys.* B **206** 321–32

[35] Haiping H and Kabashima Y 2014 Origin of the computational hardness for learning with binary synapses *Phys. Rev.* E **90** 052813

[36] Gerstner W, Kistler W M, Naud R and Paninski L 2014 *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition* (Cambridge University Press)

[37] Goel N S, Maitra S C and Montroll E W 1971 On the volterra and other nonlinear models of interacting populations *Rev. Mod. Phys.* **43** 231–76
[38] Robert M 1972 Will a large complex system be stable? *Nature* **238** 413–4
[39] Fyodorov Y V and Khoruzhenko B A 2016 Nonlinear analogue of the may-wigner instability transition *Proc. Natl Acad. Sci.* **113** 6827–32
[40] Lacroix-A-Chez-Toine B and Fyodorov Y V 2022 Counting equilibria in a random non-gradient dynamics with heterogeneous relaxation rates *J. Phys. A: Math. Theor.* **55** 144001
[41] Vivo P 2024 Random linear systems with quadratic constraints: from random matrix theory to replicas and back arXiv:2401.03209
[42] Welling M and Teh Y W 2011 Bayesian learning via stochastic gradient langevin dynamics *International Conference on Machine Learning Proceedings of the 28th International Conference on Machine Learning (ICML-11) (Bellevue, Washington, USA)* (ACM) 681–88
[43] Schuecker J, Goedeke S and Helias M 2018 Optimal sequence memory in driven random networks *Phys. Rev.* X **8** 041029
[44] Clark D G, Abbott L F and Litwin-Kumar A 2023 Dimension of activity in random neural networks *Phys. Rev. Lett.* **131** 118401
[45] Jiang Z, Chen Z, Hou T and Huang H 2023 Spectrum of non-hermitian deep-hebbian neural networks *Phys. Rev. Res.* **5** 013090
[46] Haiping H 2024 Eight challenges in developing theory of intelligence *Front. Comput. Neurosci* **18** 1388166
[47] Qiu J and Huang H 2024 (https://github.com/Qjbtiger/Equilibrium-RNN)