State Space Structure of Random Recurrent Neuronal Networks

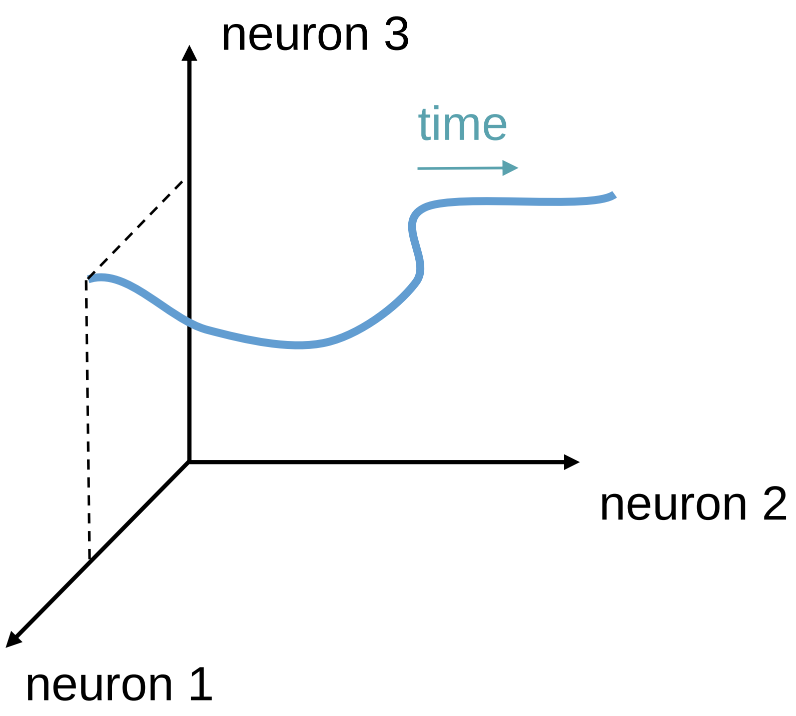
Alexander van Meegen

Current affiliation: Harvard University

Work done at: Jülich Research Centre

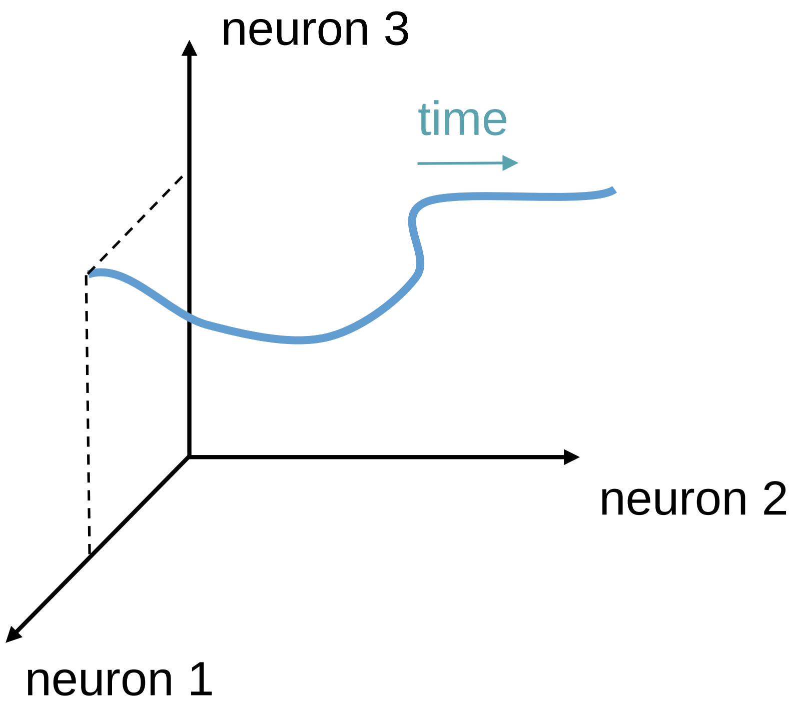
Neural Networks as Dynamical Systems

state space

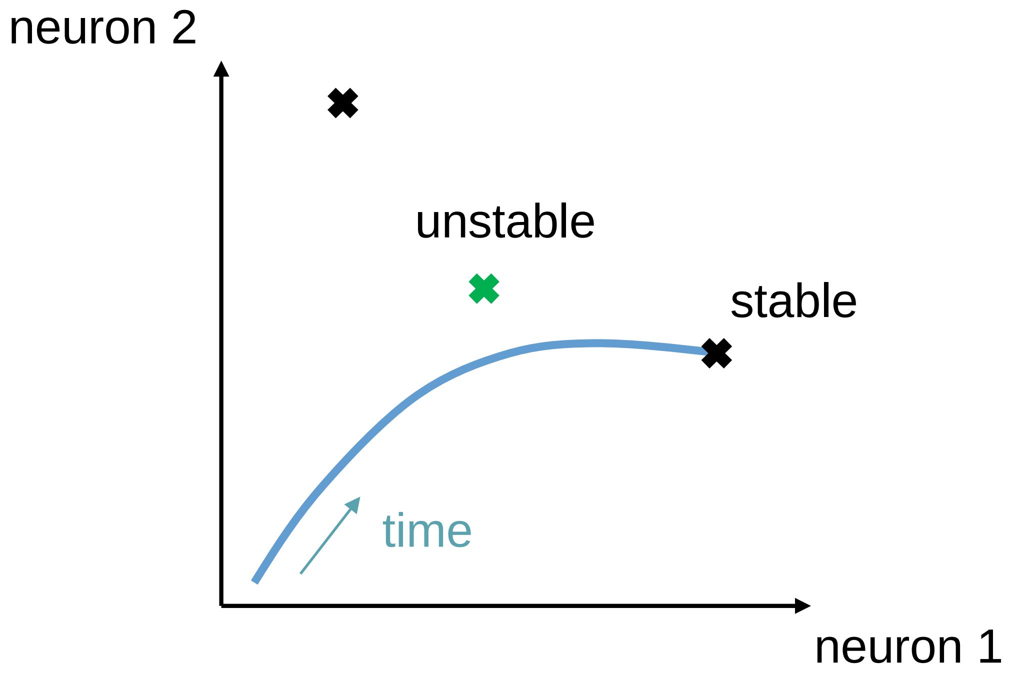


Neural Networks as Dynamical Systems

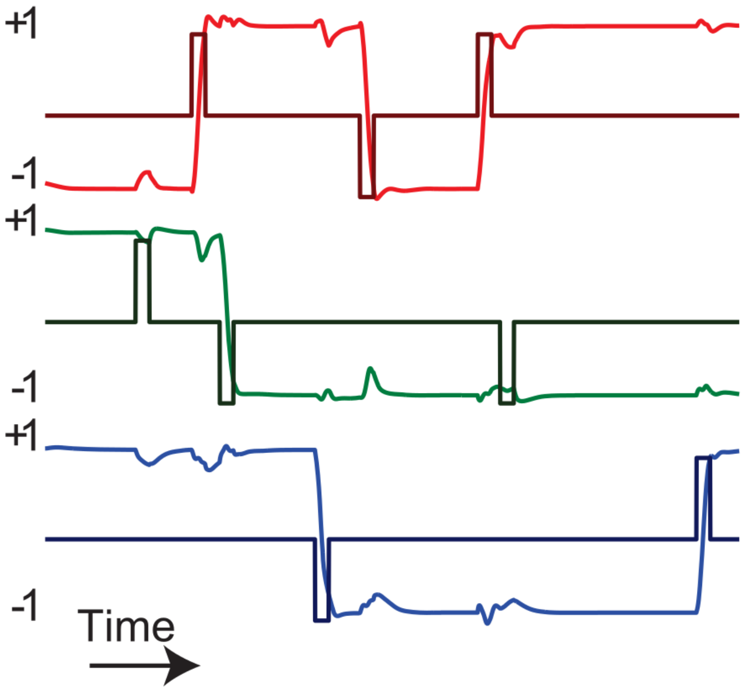
state space



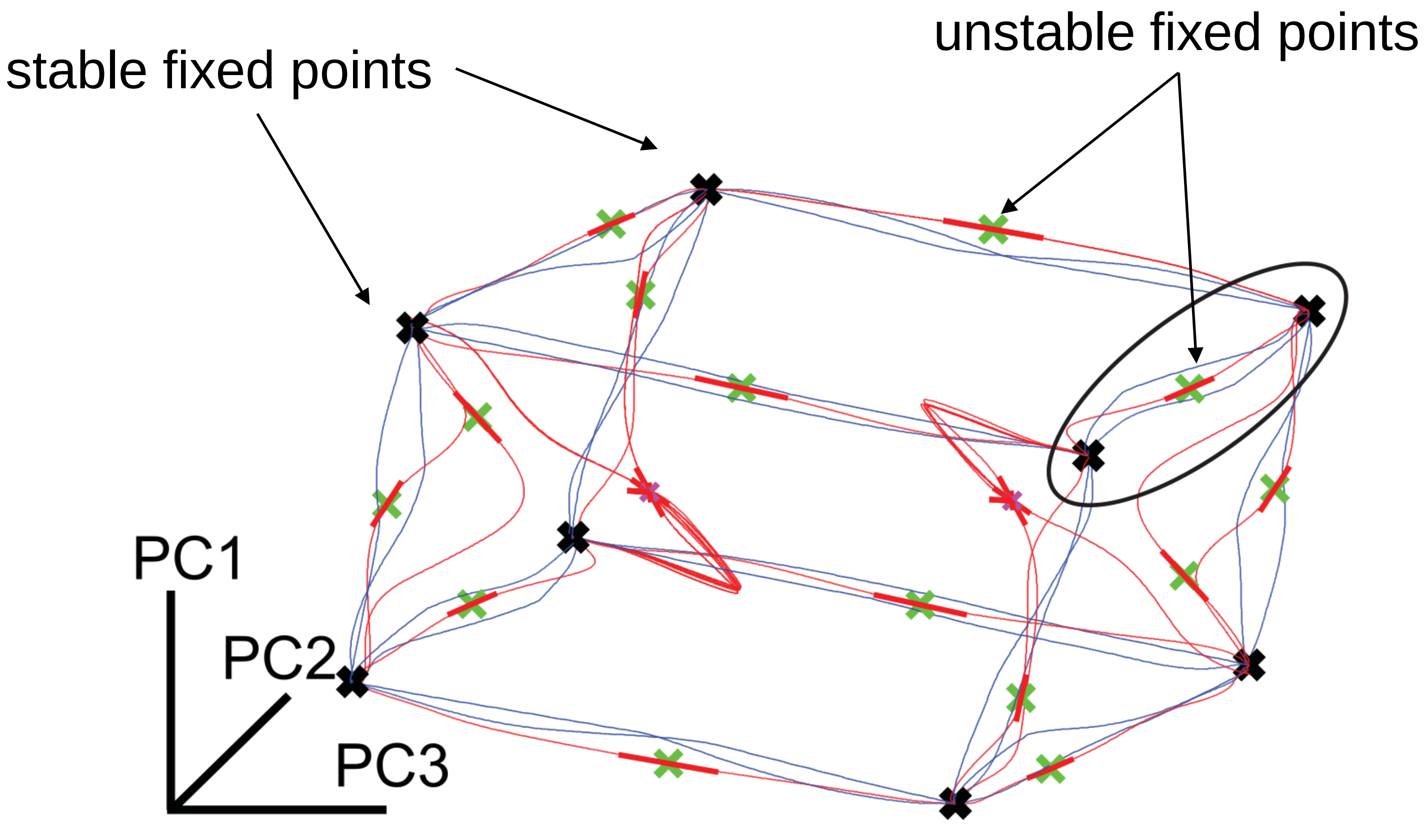
fixed points



Opening the Black Box

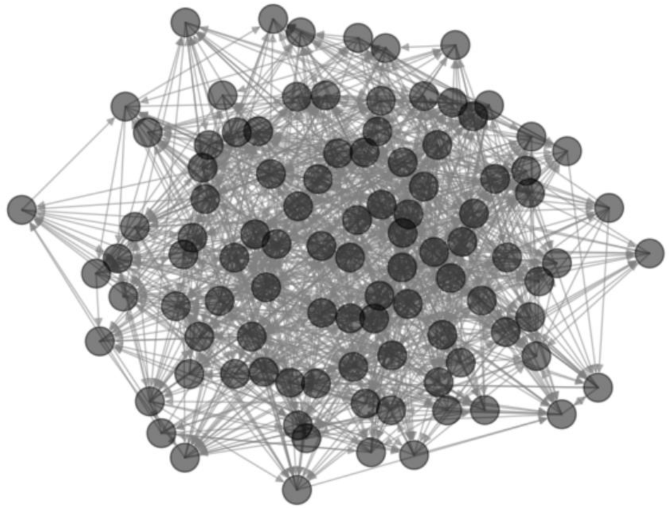


Sussillo & Barak; Neural Comput. (2013), Fig. 2



Sussillo & Barak; Neural Comput. (2013), Fig. 3

Random Recurrent Neuronal Networks

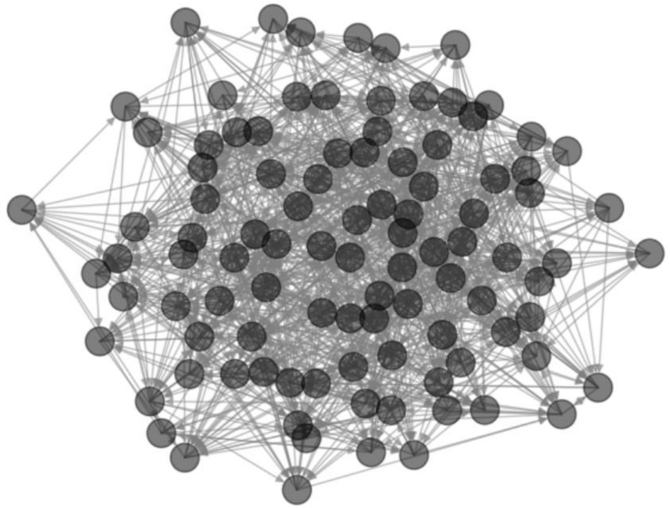


$$\boxed{\dot{x}_i = -x_i} + \boxed{\sum_{j=1}^{N} J_{ij}\phi(x_j)}$$

exponential relaxation

i.i.d. Gaussian coupling weights with strength g, tanh transfer function

Random Recurrent Neuronal Networks



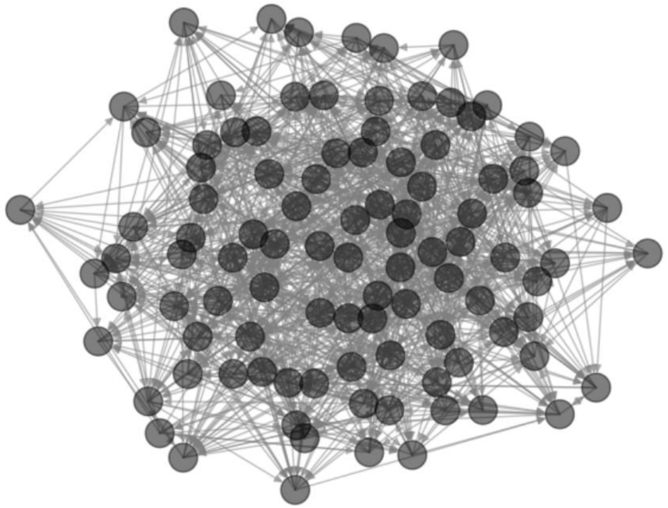
$$\boxed{\dot{x}_i = -x_i} + \boxed{\sum_{j=1}^{N} J_{ij}\phi(x_j)}$$

exponential relaxation

i.i.d. Gaussian coupling weights with strength g, tanh transfer function

- Dynamical Mean-Field Theory   Sompolinsky, Crisanti, Sommers; Phys. Rev. Lett. (1988)
  - Activity statistics: temporal autocorrelation
  - Chaos above critical connection strength g_c

Random Recurrent Neuronal Networks



$$\boxed{\dot{x}_i = -x_i} + \boxed{\sum_{j=1}^{N} J_{ij}\phi(x_j)}$$

exponential relaxation

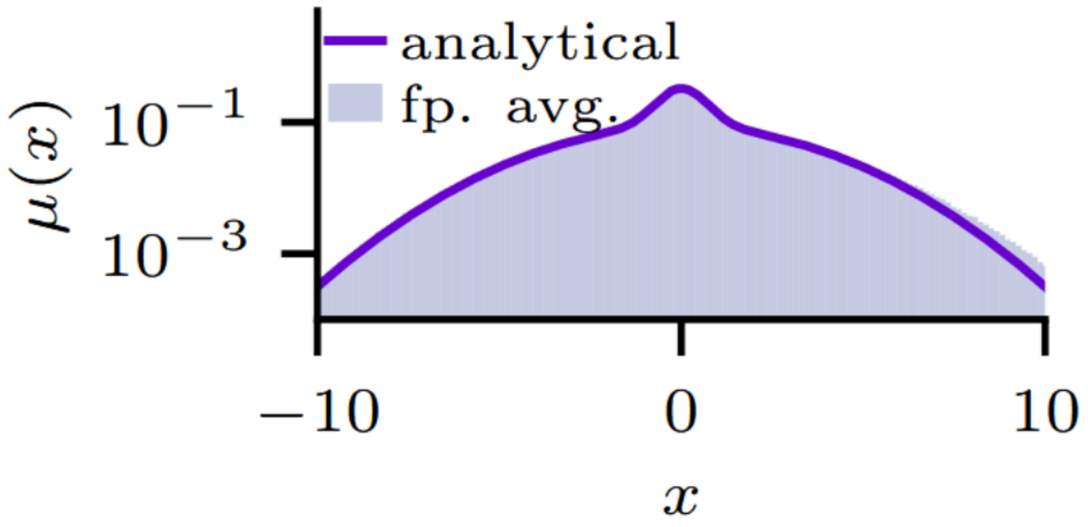
i.i.d. Gaussian coupling weights with strength g, tanh transfer function

- Dynamical Mean-Field Theory Sompolinsky, Crisanti, Sommers; Phys. Rev. Lett. (1988)
  - Activity statistics: temporal autocorrelation
  - Chaos above critical connection strength g_c
- State space & fixed points:
  - Exponential number of unstable fixed points Wainrib & Touboul; Phys. Rev. Lett. (2013)
  - *Where are the fixed points in phase space?*

Distribution of Fixed Points

- Random network → *distribution* of fixed points in state space
- Approach: Kac-Rice formula → random matrix problem

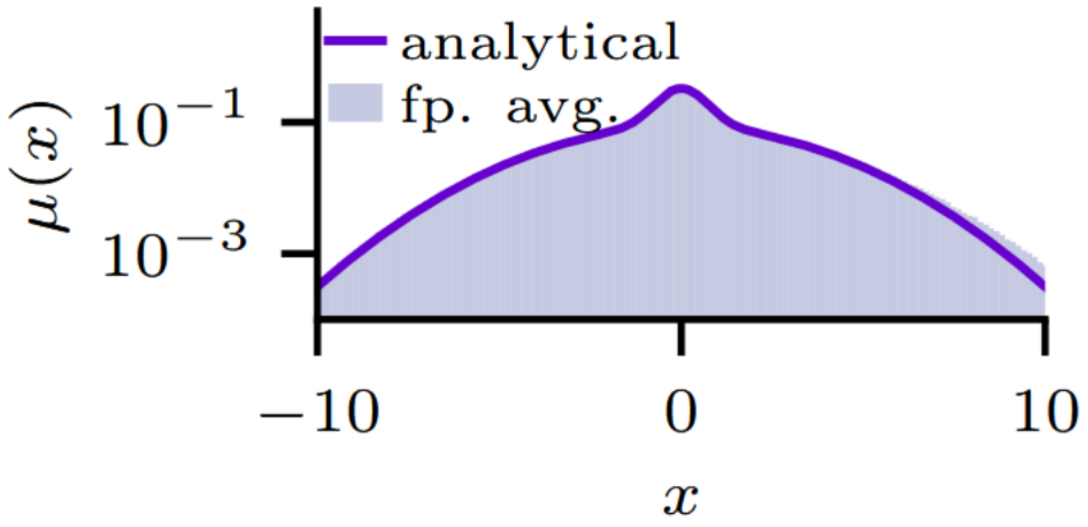
Distribution of fixed-point coordinates (of ~ 40,000 fixed points for N = 100)

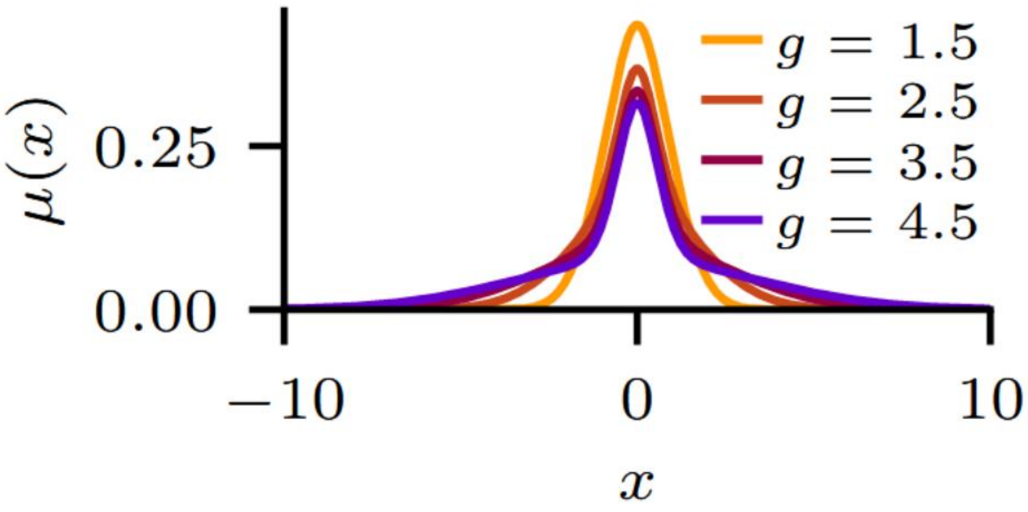


Distribution of Fixed Points

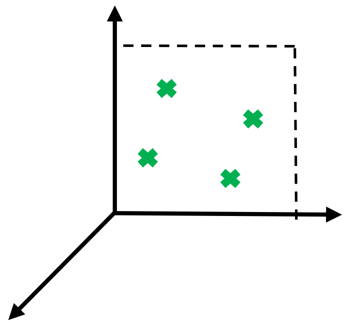
- Random network → *distribution* of fixed points in state space
- Approach: Kac-Rice formula → random matrix problem

Distribution of fixed-point coordinates (of ~ 40,000 fixed points)





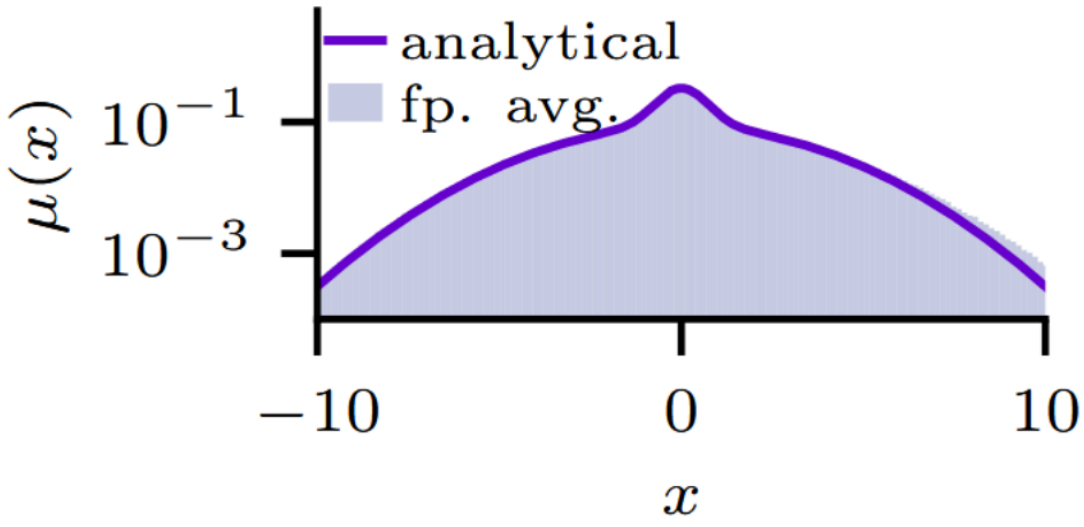
Distribution of Fixed Points

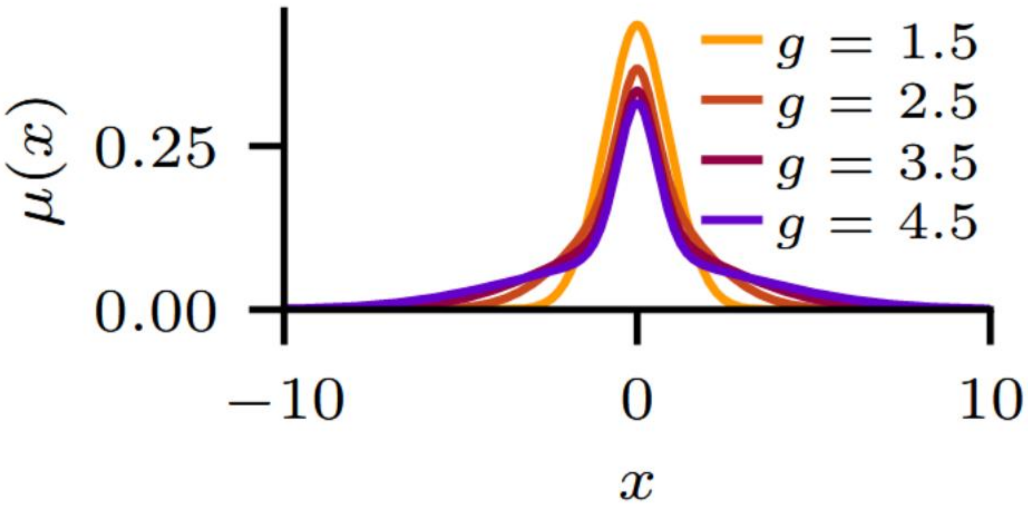


- Random network → *distribution* of fixed points in state space
- Approach: Kac-Rice formula → random matrix problem

peak at zero:
fixed points in the span
of a subset of the axes

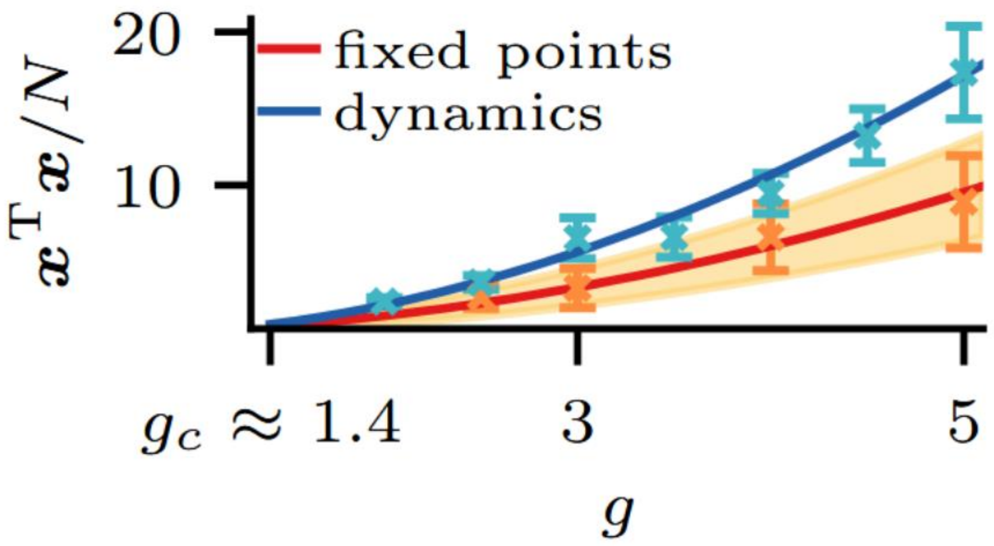
Distribution of fixed-point coordinates (of ~ 40,000 fixed points)





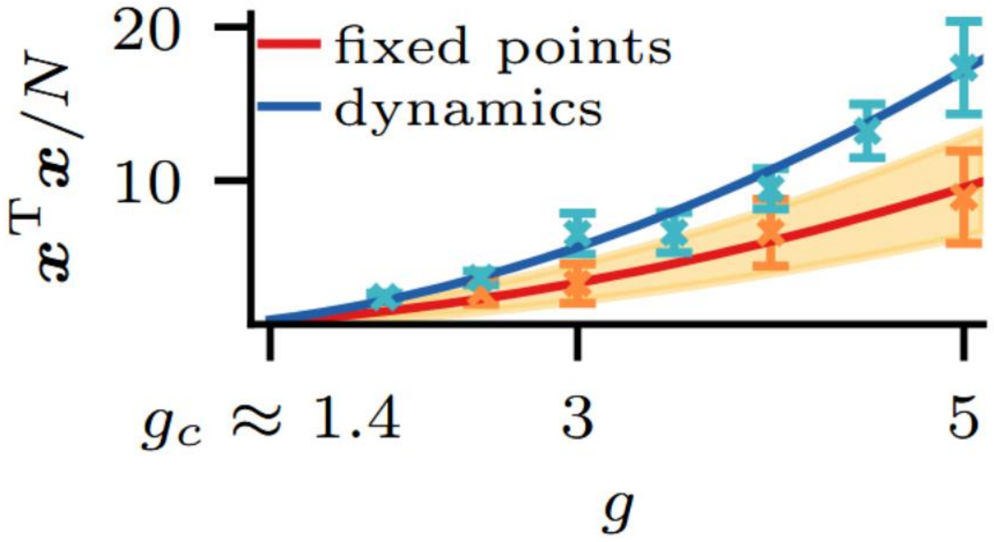
Separation of Fixed Points and Dynamics

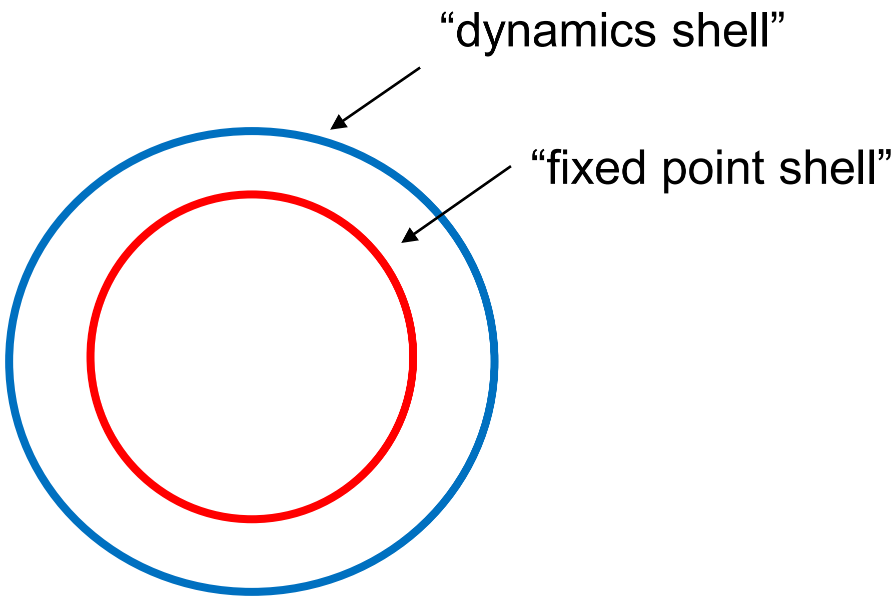
- Distance from origin in state space
  - Dynamics: Zero-lag autocorrelation from Dynamical Mean-Field Theory
  - Fixed points: From distribution of fixed points



Separation of Fixed Points and Dynamics

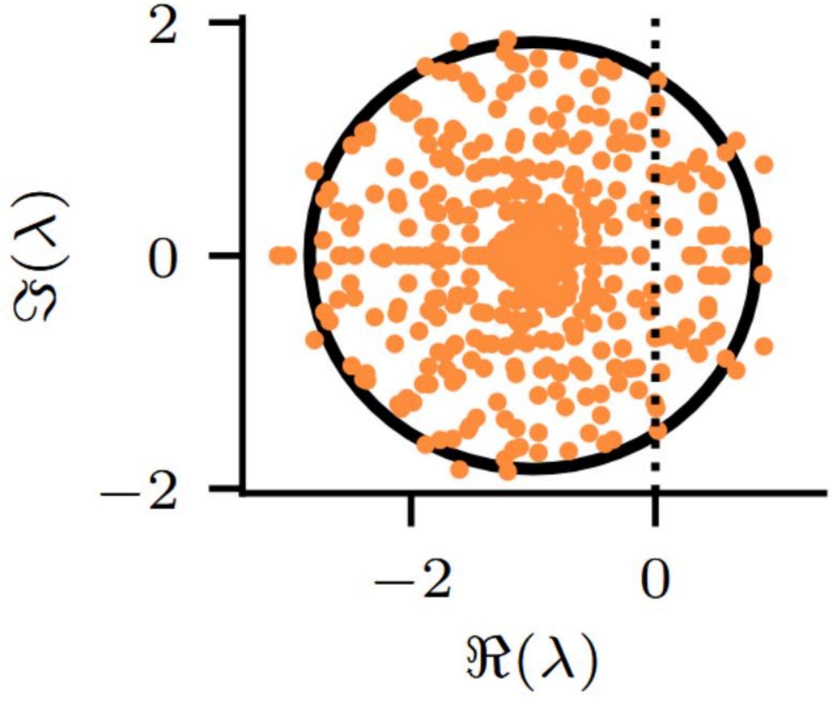
- Distance from origin in state space
  - Dynamics: Zero-lag autocorrelation from Dynamical Mean-Field Theory
  - Fixed points: From distribution of fixed points



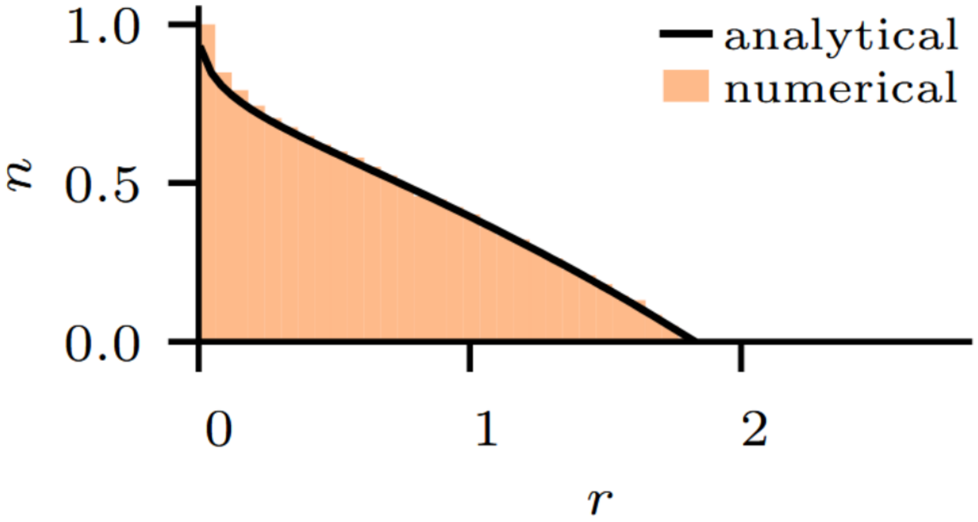


Linearized Dynamics

Jacobian spectrum at fixed points



Radial eigenvalue distribution

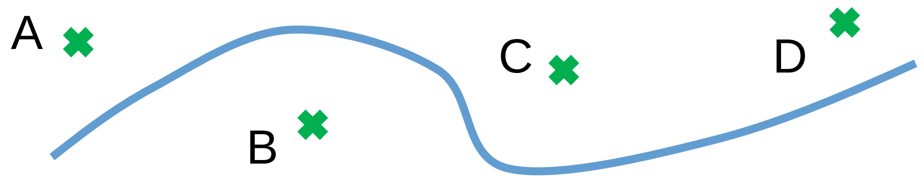


Summary & Outlook

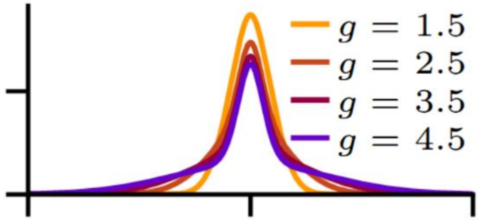
- Analytical calculation of fixed point distribution for random recurrent network
  - Located in the span of a subset of axes; radial separation in state space
  - Linearized dynamics (Jacobian)

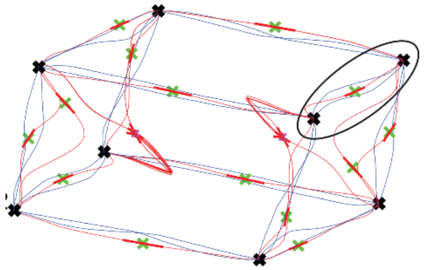
- Structural backbone for sequence processing

Rabinovich, Huerta, Laurent; Science (2008)



- How does training shape the state space structure of the network?



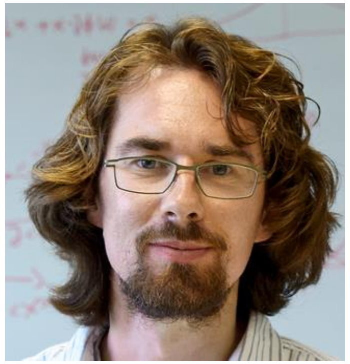


The Team

Jakob Stubenrauch

Christian Keup

Anno Kurth



Moritz Helias

Details: Stubenrauch, Keup, Kurth, Helias, van Meegen; arXiv 2210.07877 (2022)

Studienstiftung des deutschen Volkes

Kac-Rice Formula

$$\dot{\boldsymbol{x}} = \boldsymbol{y}(\boldsymbol{x}) \qquad\qquad \boldsymbol{y}(\boldsymbol{x}) = -\boldsymbol{x} + \mathbf{J}\phi(\boldsymbol{x}) + \boldsymbol{\eta}$$

Counting fixed points

$$N_{\mathrm{fp}}(V) = \int_V d\boldsymbol{x}\,\delta\left[\boldsymbol{y}(\boldsymbol{x})\right]\left|\det \boldsymbol{y}'(\boldsymbol{x})\right|$$

Kac-Rice formula

$$\rho(\boldsymbol{x}) = \left\langle \delta[\boldsymbol{y}(\boldsymbol{x})]\left|\det \boldsymbol{y}'(\boldsymbol{x})\right|\right\rangle_{\mathbf{J},\boldsymbol{\eta}}$$

$$\rho(\boldsymbol{x}) = \int d\boldsymbol{y}'\, p_{\boldsymbol{x}}(\boldsymbol{y}=0, \boldsymbol{y}')\left|\det \boldsymbol{y}'\right|$$

Velocity & Jacobian Statistics

$$\boldsymbol{y}(\boldsymbol{x}) = -\boldsymbol{x} + \mathbf{J}\phi(\boldsymbol{x}) + \boldsymbol{\eta} \qquad\qquad \boldsymbol{y}'(\boldsymbol{x}) = -\mathbb{1} + \mathbf{J}\,\mathrm{diag}\left[\phi'(\boldsymbol{x})\right]$$

$$\langle y_i(\boldsymbol{x})\rangle = -x_i \equiv \mu_i(\boldsymbol{x}), \qquad \langle [\boldsymbol{y}'(\boldsymbol{x})]_{ik}\rangle = -\delta_{ik} \equiv [\boldsymbol{\mu}_i(\boldsymbol{x})]_k$$

$$\langle\langle y_i(\boldsymbol{x}) y_j(\boldsymbol{x})\rangle\rangle = \delta_{ij}\left[\frac{g^2}{N}\sum_k \phi(x_k)^2 + D\right] \equiv \delta_{ij}[\kappa(\boldsymbol{x}) + D]$$

$$\left\langle\left\langle y_i(\boldsymbol{x})\,[\boldsymbol{y}'(\boldsymbol{x})]_{jk}\right\rangle\right\rangle = \frac{g^2}{N}\delta_{ij}\phi(x_k)\phi'(x_k) \equiv \delta_{ij}[\boldsymbol{k}(\boldsymbol{x})]_k,$$

$$\left\langle\left\langle [\boldsymbol{y}'(\boldsymbol{x})]_{ik}\,[\boldsymbol{y}'(\boldsymbol{x})]_{jl}\right\rangle\right\rangle = \delta_{ij}\delta_{kl}\frac{g^2}{N}\phi'(x_k)^2 \equiv \delta_{ij}[\mathbf{K}(\boldsymbol{x})]_{kl}.$$

Random Matrix Problem

Kac-Rice formula

$$\rho(\boldsymbol{x}) = \int d\boldsymbol{y}'\, p_{\boldsymbol{x}}(\boldsymbol{y}=0, \boldsymbol{y}')\ |\det \boldsymbol{y}'|$$

$$\rho(\boldsymbol{x}) = p_{\mathrm{L}}(\boldsymbol{x})\ \langle |\det \boldsymbol{y}'| \rangle_{\boldsymbol{y}' \sim p_{\boldsymbol{x}}(\boldsymbol{y}'|\boldsymbol{y}=0)}$$

$$\rho(\boldsymbol{x}) = p_{\mathrm{L}}(\boldsymbol{x})\ \langle |\det [\mathbf{M}(\boldsymbol{x}) + \mathbf{X}\,\boldsymbol{\Sigma}(\boldsymbol{x})]| \rangle_{X_{ij} \overset{\mathrm{i.i.d.}}{\sim} \mathcal{N}(0,1/N)}$$

Fixed Point Distribution

$$\rho(\boldsymbol{x}) \doteq \exp\Big(- N S(\boldsymbol{x}) \Big)$$

$$S(\boldsymbol{x}) = \frac{q(\boldsymbol{x})}{2[\kappa(\boldsymbol{x}) + D]} + \frac{1}{2}\ln\{2\pi[\kappa(\boldsymbol{x}) + D]\} - \zeta(\boldsymbol{x})$$

$$q(\boldsymbol{x}) = \tfrac{1}{N}\textstyle\sum_{i=1}^{N} x_i^2 \qquad\qquad \kappa(\boldsymbol{x}) = \tfrac{g^2}{N}\textstyle\sum_{i=1}^{N} \phi(x_i)^2$$

$$\zeta(\boldsymbol{x}) = -\frac{1}{2} z_* + \frac{1}{2N}\sum_{i=1}^{N} \ln[1 + z_* g^2 \phi'(x_i)^2] \qquad\qquad 1 = \frac{1}{N}\sum_{i=1}^{N} \frac{g^2\phi'(x_i)^2}{1 + z_*\, g^2\phi'(x_i)^2}$$

Distribution of Coordinates

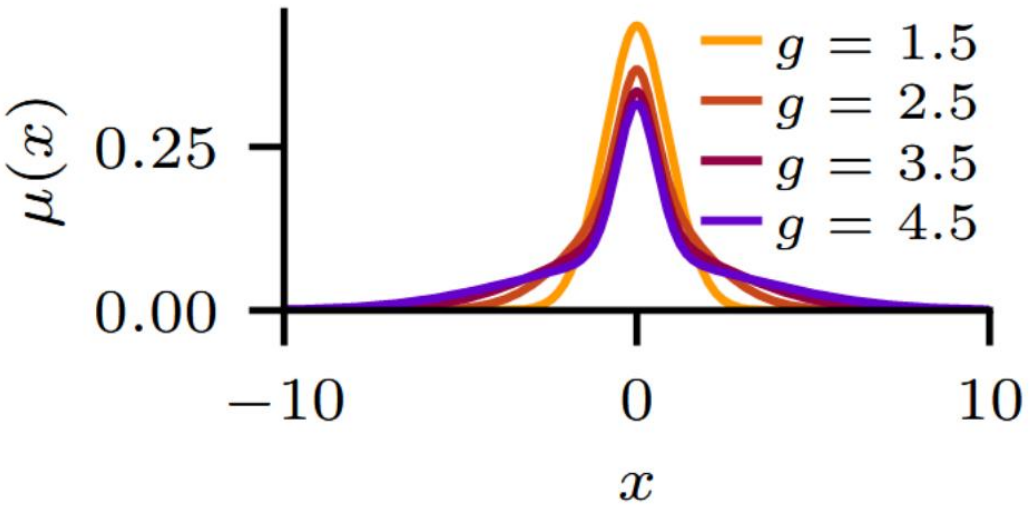
$$\mu_{\boldsymbol{x}}(y) = \frac{1}{N}\sum_{i=1}^{N}\delta(y - x_i)$$

$$\mu_*(y) \propto \sqrt{1 + \alpha\phi'(y)^2}\, e^{-\frac{y^2}{2\beta} + \gamma\phi(y)^2}$$

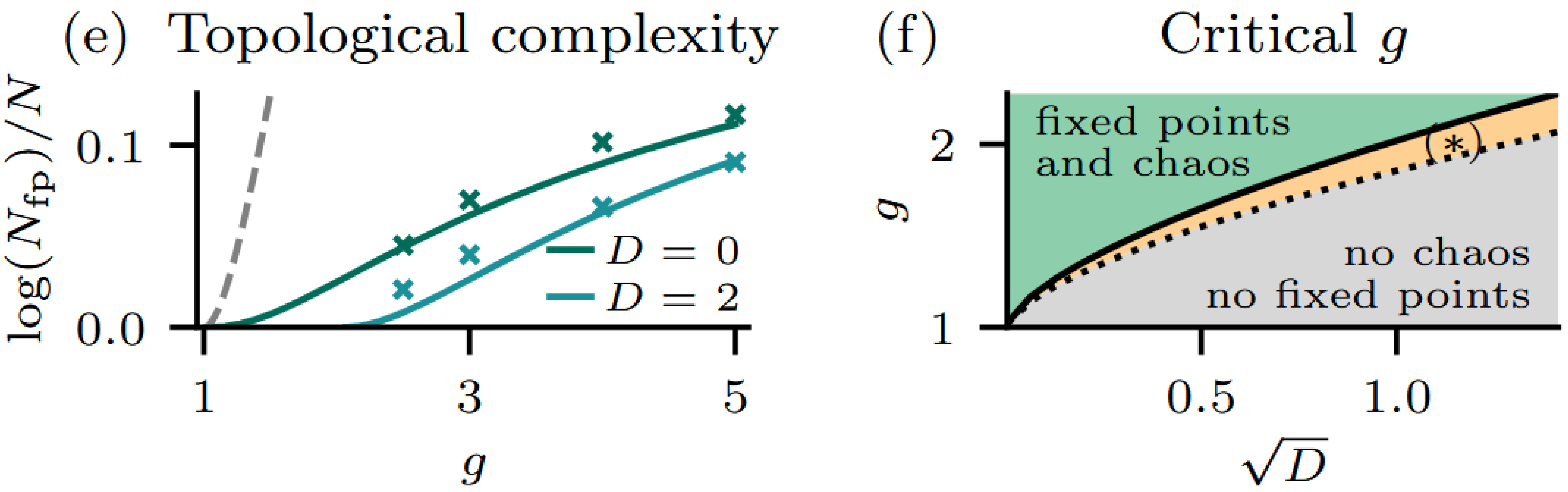
$$1 = g^2 \left\langle \left(\phi'(y)^{-2} + \alpha\right)^{-1} \right\rangle_{\mu_*}$$

$$\beta = \left\langle \phi(y)^2 \right\rangle_{\mu_*} + D$$

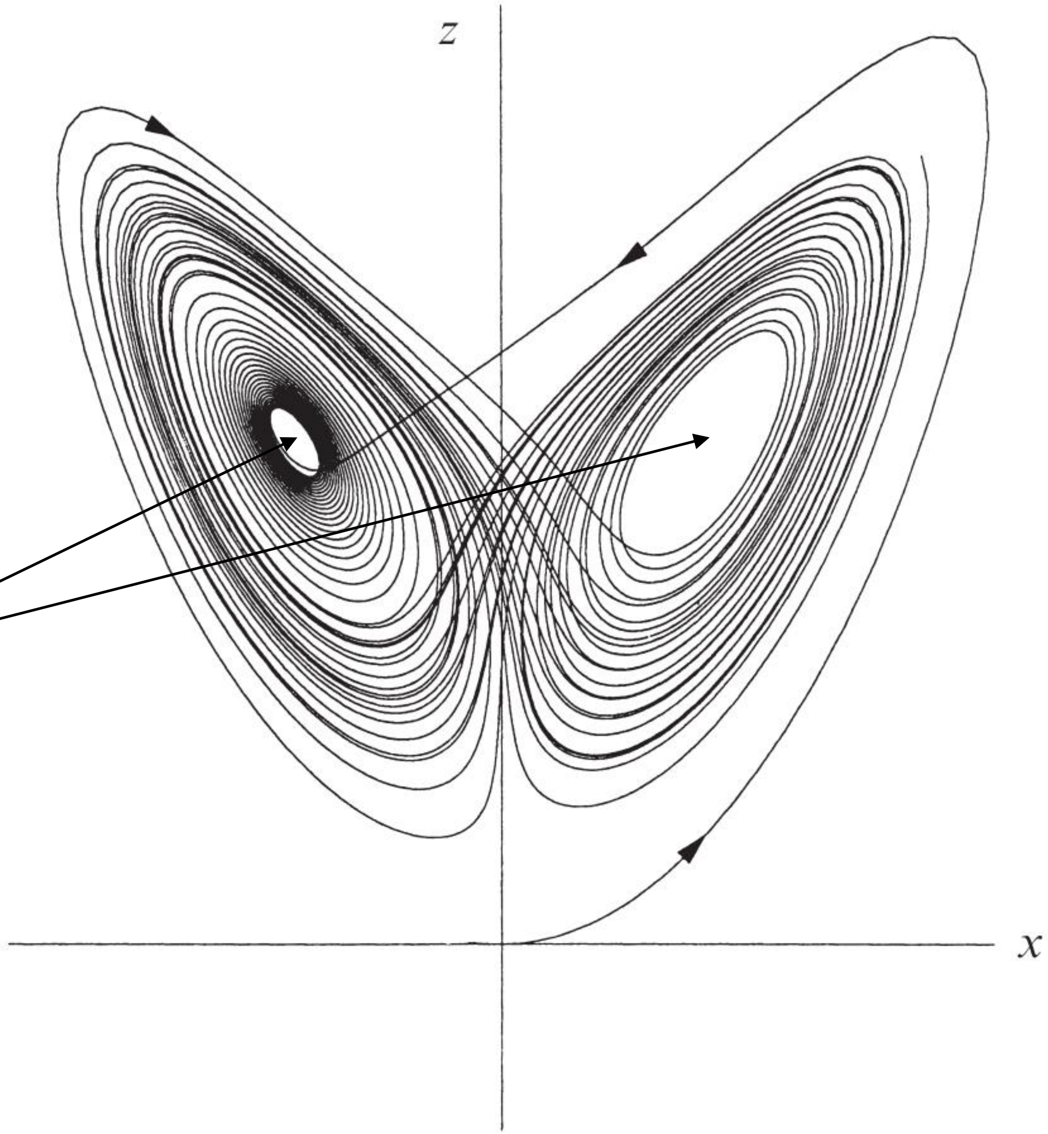
$$\gamma = \frac{g^2}{2\beta}\left(\beta^{-1}\left\langle y^2 \right\rangle_{\mu_*} - 1\right)$$



Number of Fixed Points



Lorenz Attractor



unstable fixed points

Strogatz – Nonlinear Dynamics and Chaos, Fig. 9.3.2