

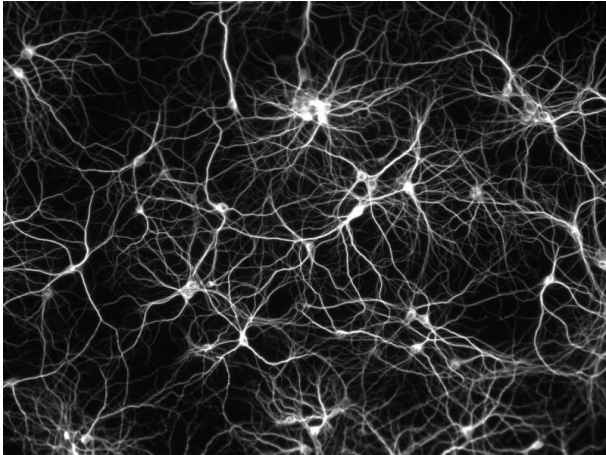
Neural Networks Dynamics

Bruno Cessac

NeuroMathComp Team, INRIA Sophia Antipolis, France.

20-01-14

Neural networks



- Multi-scale spatio-temporal non-linear dynamics.

Neural networks

- Multi-scale spatio-temporal non-linear dynamics.
- Neural networks share fascinating capacities (adaptation, memory, generalisation capabilities, ...).

Neural networks

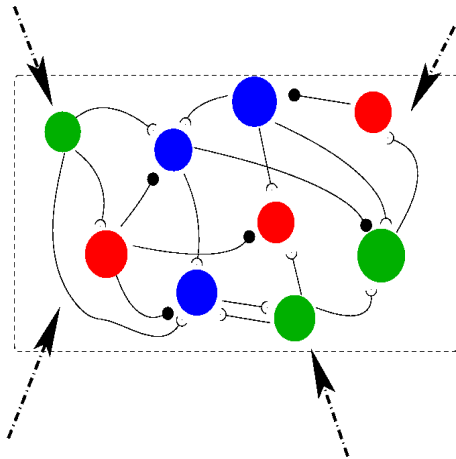
- Multi-scale spatio-temporal non-linear dynamics.
- Neural networks share fascinating capacities (adaptation, memory, generalisation capabilities, ...).
- Understanding these capacities via (mathematical) models.

Neural networks

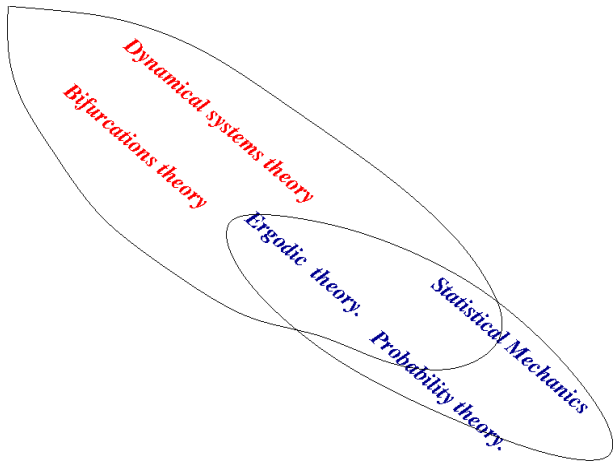
- Multi-scale spatio-temporal non-linear dynamics.
- Neural networks share fascinating capacities (adaptation, memory, generalisation capabilities, ...).
- Understanding these capacities via (mathematical) models.
 - A model is a set of equations attempting to mimic some aspects of neural dynamics.
 - We don't want to include all bio-physical details.
 - We want a trade-off between a realistic description and mathematical/numerical tractability.
 - A model \Leftrightarrow a set of well posed questions.

At the current state of knowledge we need new mathematical concepts and methods for a proper analysis of the brain dynamics.

Neural networks



- Multi-scale spatio-temporal dynamics.
- A neural network is formally a graph where the nodes are the neurons and the edges are the synapses.
- Dynamics results from a complex interplay between neurons dynamics and network structure, which is itself evolving.
- Analysing models with theoretical tools such as dynamical systems theory and statistical physics.
- Extracting mechanisms and principles from this analysis and apply/compare to artificial/biological neural networks.



We are going to consider a single example of neural network (Simple but Non Trivial Model) (SNT).

We shall address mathematically the following questions.

We are going to consider a single example of neural network (Simple but Non Trivial Model) (SNT).

We shall address mathematically the following questions.

- Dynamical analysis and generic dynamics.
- Mean-field methods: presentation, justification and extensions.
- Response to stimuli.
- Is there a natural notion of connectivity in a neural network ?
- Plasticity (synaptic and intrinsic).

The model

The model

The Amari-Wilson-Cowan model (amari:72,wilson-cowan:72):

$$\frac{du_i}{dt} = -\frac{u_i}{\tau} + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i; \quad (1)$$

- u_i Membrane potential of neuron i .
- τ leak rate time constant.
- J_{ij} synaptic weight from j (pre-synaptic neuron) to i (post-synaptic neuron).
- f is a sigmoidal function ($\tanh(gx)$; $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{gx} e^{-\frac{t^2}{2}} dt$).
- θ_i : external stimulus imposed upon neuron i .

The model

$$\frac{du_i}{dt} = -\frac{u_i}{\tau} + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i.$$

The model

$$\frac{du_i}{dt} = -\frac{u_i}{\tau} + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i.$$

- Time discretization. $\frac{du_i}{dt} \rightarrow \frac{u_i(t+dt) - u_i(t)}{dt}$. $dt \ll \tau$

The model

$$\frac{du_i}{dt} = -\frac{u_i}{\tau} + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i.$$

- Time discretization. $\frac{du_i}{dt} \rightarrow \frac{u_i(t+dt) - u_i(t)}{dt}$. $dt \ll \tau$
- $\gamma = 1 - \frac{dt}{\tau}$; $\gamma \in [0, 1[$

The model

$$\frac{du_i}{dt} = -\frac{u_i}{\tau} + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i.$$

- Time discretization. $\frac{du_i}{dt} \rightarrow \frac{u_i(t+dt) - u_i(t)}{dt}$. $dt \ll \tau$
- $\gamma = 1 - \frac{dt}{\tau}$; $\gamma \in [0, 1[$
- $dt = 1$.

The model

$$\frac{du_i}{dt} = -\frac{u_i}{\tau} + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i.$$

- Time discretization. $\frac{du_i}{dt} \rightarrow \frac{u_i(t+dt) - u_i(t)}{dt}$. $dt \ll \tau$
- $\gamma = 1 - \frac{dt}{\tau}$; $\gamma \in [0, 1[$
- $dt = 1$.
- Add Gaussian noise $\sigma B_i(t)$. $B_i(t)$ i.i.d.r.v. $\mathcal{N}(0, 1)$.

$$u_i(t+1) = \gamma u_i(t) + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i + \sigma B_i(t). \quad (2)$$

The SNT model

$$u_i(t + 1) = \gamma u_i(t) + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i + \sigma B_i(t).$$

- N -dimensional stochastic non linear dynamical system;
- Depending on $N^2 + N + 3$ parameters (N^2 synaptic weights, N stimuli, the gain g , the leak γ , the noise intensity σ);
- Later on we shall consider the case where the stimulus θ_i depends on time (linear response);
- We shall also consider the case where J_{ij} depend on time via the network history (synaptic plasticity).

Dynamical analysis and generic dynamics

$$u_i(t+1) = \gamma u_i(t) + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i + \sigma B_i(t).$$

$$\mathbf{F}(\mathbf{u}) = \gamma \mathbf{u} + \mathcal{J}.f(\mathbf{u}) - \boldsymbol{\theta} \quad (3)$$

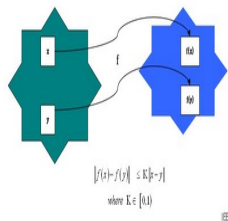
$$F_i(\mathbf{u}) = \gamma u_i + \sum_{j=1}^N J_{ij} f(u_j) - \theta_i$$

$$\mathbf{u}(t+1) = \mathbf{F}(\mathbf{u}(t)) + \sigma \mathbf{B}(t). \quad (4)$$

Contraction

Theorem

$\forall \gamma, \theta, \mathcal{J}, N$ there is a g value, $g_{as} > 0$ such that for $0 \leq g < g_{as}$, \mathbf{F} is a contraction.



$$D\mathbf{F}_{\mathbf{u}} = \gamma\mathcal{I} + g\mathcal{J}.\Lambda(\mathbf{u}), \quad (5)$$

$$\Lambda_{ij}(\mathbf{u}) = f'_g(u_i)\delta_{ij}$$

Contraction

Corollary

$\forall \gamma, \theta, \mathcal{J}, N$ there is a g value, $g_{as} > 0$ such that for $0 \leq g < g_{as}$ the dynamical system (2) has a unique stable equilibrium attracting all trajectories (for $\sigma = 0$).

Neurons membrane potential converge to a constant value.

Example

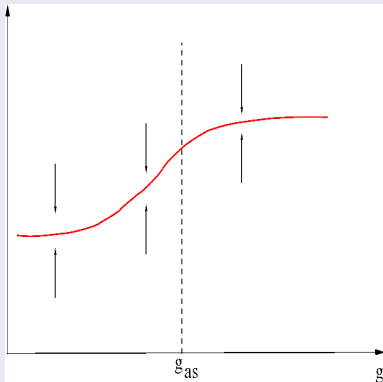
$$f(x) = \tanh(gx); \theta_i = 0$$

\Rightarrow

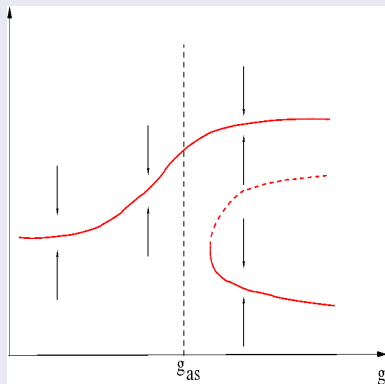
$u_i = 0$ is a equilibrium point.

Contraction

One equilibrium



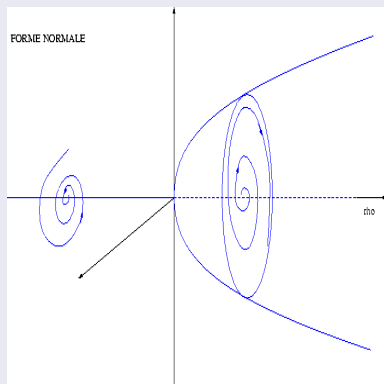
Saddle-node bifurcation



Coexistence of several equilibrium.

Dynamical bifurcations

Hopf bifurcation



Appearance of oscillations in dynamics.

Remark

All these bifurcations are codimension one bifurcations for which a general classification exist.

For codimensions > 2 (especially $N^2 + N + 3$) this is "terra incognita".

Sampling the parameters space with random parameters

Random sampling

Random sampling

- N^2 synaptic weights \Rightarrow sampled by a Gaussian distribution controlled by two parameters: mean and covariance $\mathbb{E}[J_{ij}]$, $\text{Cov}[J_{ij}, J_{kl}]$.

Sampling the parameters space with random parameters

Random sampling

- N^2 synaptic weights \Rightarrow sampled by a Gaussian distribution controlled by two parameters: mean and covariance $\mathbb{E}[J_{ij}]$, $\text{Cov}[J_{ij}, J_{kl}]$.

Example 1: fully connected, i.i.d., one population. Amari, 72; Sompolinsky et al., 1988; ...

$$\mathbb{E}[J_{ij}] = \bar{J}, \quad \text{Cov}[J_{ij}, J_{kl}] = \frac{J^2}{N} \delta_{i,j} \delta_{k,l}$$

Sampling the parameters space with random parameters

Random sampling

- N^2 synaptic weights \Rightarrow sampled by a Gaussian distribution controlled by two parameters: mean and covariance $\mathbb{E}[J_{ij}]$, $\text{Cov}[J_{ij}, J_{kl}]$.

Example 2: fully connected, i.i.d. several populations. e.g. Faugeras et al, 2008

$$\mathbb{E}[J_{ij}] = \frac{\bar{J}_{\alpha,\beta}}{N_\beta}, \quad \text{Cov}[J_{ij}, J_{kl}] = \frac{J_{\alpha,\beta}^2}{N_\beta} \delta_{i,j} \delta_{k,l}$$

Random sampling

- N^2 synaptic weights \Rightarrow sampled by a Gaussian distribution controlled by two parameters: mean and covariance $\mathbb{E}[J_{ij}]$, $\text{Cov}[J_{ij}, J_{kl}]$.

Example 3: fully connected, correlated. Faugeras & MacLaurin, 2013

$$\mathbb{E}[J_{ij}] = \frac{\bar{J}}{N}, \quad \text{Cov}[J_{ij}, J_{kl}] = \frac{C_{ij,kl}}{N}$$

Random sampling

- N^2 synaptic weights \Rightarrow sampled by a Gaussian distribution controlled by two parameters: mean and covariance $\mathbb{E}[J_{ij}]$, $\text{Cov}[J_{ij}, J_{kl}]$.

Example 4: Sparse; and so on.

Random sampling

- N^2 synaptic weights \Rightarrow sampled by a Gaussian distribution controlled by two parameters: mean and covariance $\mathbb{E}[J_{ij}]$, $\text{Cov}[J_{ij}, J_{kl}]$.

Example 1: fully connected, i.i.d., one population. Amari, 72; Sompolinsky et al., 1988; ...

$$\mathbb{E}[J_{ij}] = \bar{J}, \quad \text{Cov}[J_{ij}, J_{kl}] = \frac{J^2}{N} \delta_{i,j} \delta_{k,l}$$

Sampling the parameters space with random parameters

Random sampling

- N^2 synaptic weights \Rightarrow sampled by a Gaussian distribution controlled by two parameters: mean and covariance $\mathbb{E}[J_{ij}]$, $\text{Cov}[J_{ij}, J_{kl}]$.

Example 1: fully connected, i.i.d., one population. Amari, 72; Sompolinsky et al., 1988; ...

$$\mathbb{E}[J_{ij}] = \bar{J}, \quad \text{Cov}[J_{ij}, J_{kl}] = \frac{J^2}{N} \delta_{i,j} \delta_{k,l}$$

- N stimuli \Rightarrow sampled by a Gaussian distribution controlled by two parameters: mean and covariance $\mathbb{E}[\theta_i]$, $\text{Cov}[\theta_i, \theta_j]$.

Fully connected, i.i.d., one population.

$$\mathbb{E}[\theta_i] = \bar{\theta}; \quad \text{Var}[\theta_i] = \sigma_{\theta}^2$$

The SNT model

$$u_i(t+1) = \gamma u_i(t) + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i + \sigma B_i(t).$$

We have now a N -dimensional, non-linear dynamical system + noise, where parameters are additionally random (and *quenched*).

The SNT model

$$u_i(t+1) = \gamma u_i(t) + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i + \sigma B_i(t).$$

We have now a N -dimensional, non-linear dynamical system + noise, where parameters are additionally random (and *quenched*).

POURQUOI,
POURQUOI,
POURQUOI ?



Random matrices spectra and bifurcation analysis

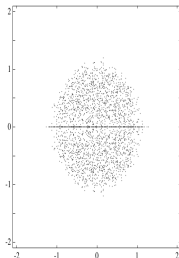
$$\mathcal{D}\mathbf{F}_{\mathbf{u}} = \gamma\mathcal{I} + g\mathcal{J}.\Lambda(\mathbf{u}), \quad \Lambda_{ij}(\mathbf{u}) = f'_g(u_i)\delta_{ij}$$

If \mathbf{u}^* is a fixed point, its stability is given by the spectral radius $\rho(\mathcal{D}\mathbf{F}_{\mathbf{u}^*})$.

The spectrum of $\mathcal{D}\mathbf{F}_{\mathbf{u}^*}$ is random because here \mathcal{J} is random.

The spectrum of \mathcal{J} is known from random matrices theory (in the balanced case).

(Girko, 1985; Edelman, 1988; Sompolinsky et al. 1988).



Generic scenario as g grows (no noise)

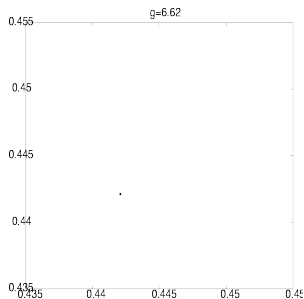
$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

Figures from E. Daucé.

Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

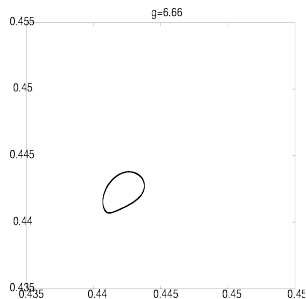
Figures from E. Daucé.



Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

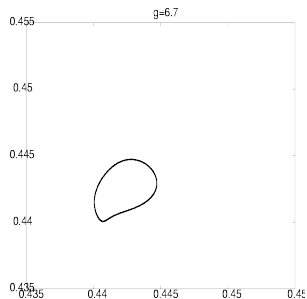
Figures from E. Daucé.



Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

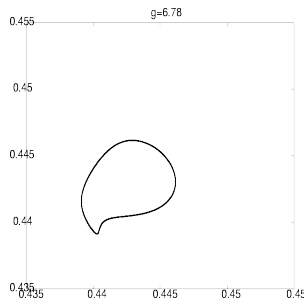
Figures from E. Daucé.



Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

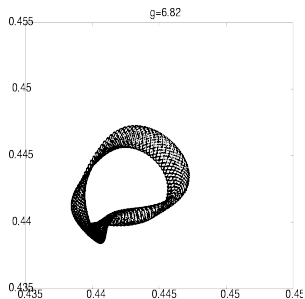
Figures from E. Daucé.



Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

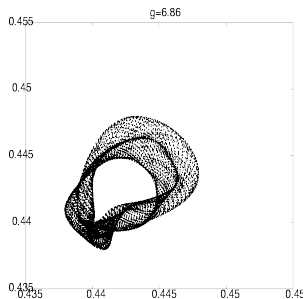
Figures from E. Daucé.



Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

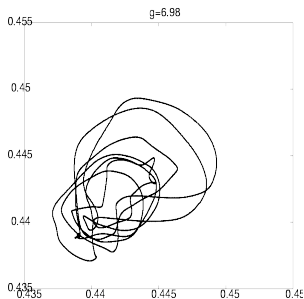
Figures from E. Daucé.



Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

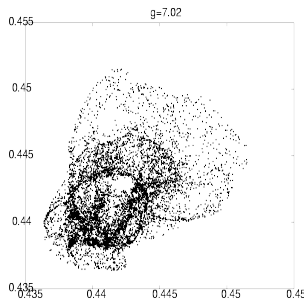
Figures from E. Dauce.



Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

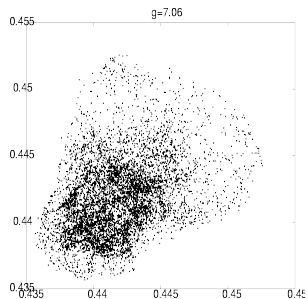
Figures from E. Daucé.



Generic scenario as g grows (no noise)

$$m(t) = \frac{1}{N} \sum_{j=1}^N f(u_j(t))$$

Figures from E. Daudé.



Generic scenario as g grows (no noise)

This is the Ruelle-Takens-Newhouse transition to chaos by quasiperiodicity

Ruelle-Takens, 1971, Newhouse, Ruelle, Takens, 1973.

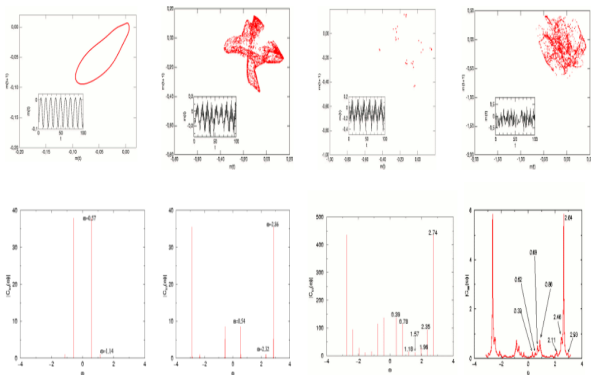


Figure: From Cessac, Sepulchre, 2006.

Maximal Lyapunov exponent

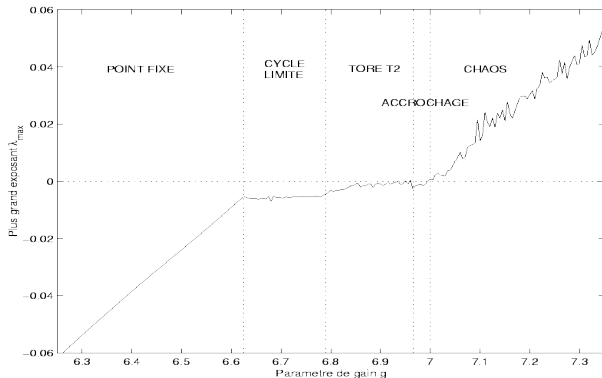
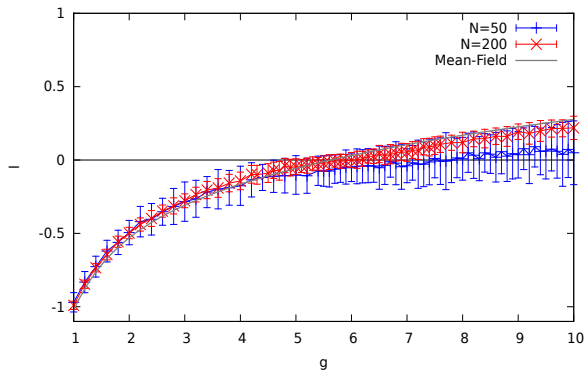


Figure: Figures from E. Daucé.

Maximal Lyapunov exponent



Mean-field analysis and bifurcation map

3 ways to obtain Mean-Field equations

3 ways to obtain Mean-Field equations

- (i) The local chaos hypothesis (Amari, 1972; Doyon et al, 1994; Cessac, 95).

3 ways to obtain Mean-Field equations

- (i) The local chaos hypothesis (Amari, 1972; Doyon et al, 1994; Cessac, 95).
- (ii) The generating functional approach developed for spin-glasses and used, for the first time, by Sompolinsky and coworkers (1988), for the continuous time dynamics, and Molgedey et al (1992) for the model (2)

3 ways to obtain Mean-Field equations

- (i) The local chaos hypothesis (Amari, 1972; Doyon et al, 1994; Cessac, 95).
- (ii) The generating functional approach developed for spin-glasses and used, for the first time, by Sompolinsky and coworkers (1988), for the continuous time dynamics, and Molgedey et al (1992) for the model (2)
- (iii) The large deviations technique introduced by Guionnet and Ben Arous for spin-glasses in 1995 and used by Moynot and Samuelides in 2002. This technique extends to correlated synaptic weights (Faugeras et MacLaurin, 2013; Muratori et al, in preparation).

Mean-field analysis and bifurcation map

$$u_i(t+1) = \gamma u_i(t) + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i + \sigma B_i(t).$$

$$G_i(t) = \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i + \sigma B_i(t).$$

(local neuronal field)

Thermodynamic limit $N \rightarrow \infty$

$G_i(t)$ converges (in law) to a Gaussian process. As a consequence, $u_i(t)$ converges to a Gaussian process with mean $\mu_i(t)$ and covariance $\Delta_{ij}(t, s)$.

Dynamic mean-field equations

A closed set of equations for the mean and covariance of the Gaussian process $u_i(t)$.

Dynamic mean-field equations

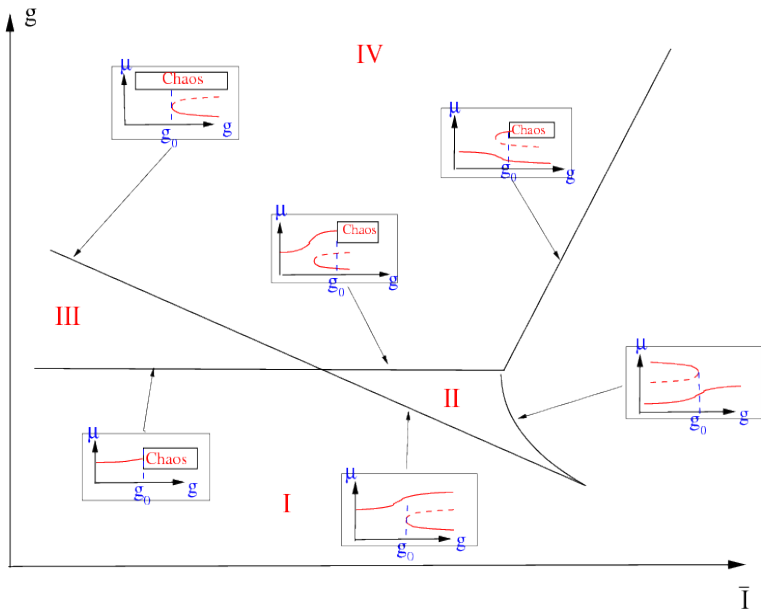
A closed set of equations for the mean and covariance of the Gaussian process $u_i(t)$.

Example 1: i.i.d. synaptic weights (Cessac, 1995).

$$\begin{aligned}\mu(t+1) &= \bar{J} \int_{\mathbb{R}} f\left(\sqrt{v(t)}h + \mu(t)\right) Dh - \bar{\theta}, \\ v(t+1) &= J^2 \int_{\mathbb{R}} f\left(\sqrt{v(t)}h + \mu(t)\right)^2 Dh + \sigma_{\theta}^2 + \sigma^2, \\ \Delta(t+1, s+1) &= J^2 \int_{\mathbb{R}^2} f\left(\frac{\sqrt{v(t)v(s)-\Delta^2(t,s)}}{\sqrt{v(s)}}h + \frac{\Delta(t,s)}{\sqrt{v(s)}}h' + \mu(t)\right) \\ &\quad f\left(\sqrt{v(s)}h' + \mu(s)\right) DhDh' + \sigma_{\theta}^2 + \sigma^2,\end{aligned}$$

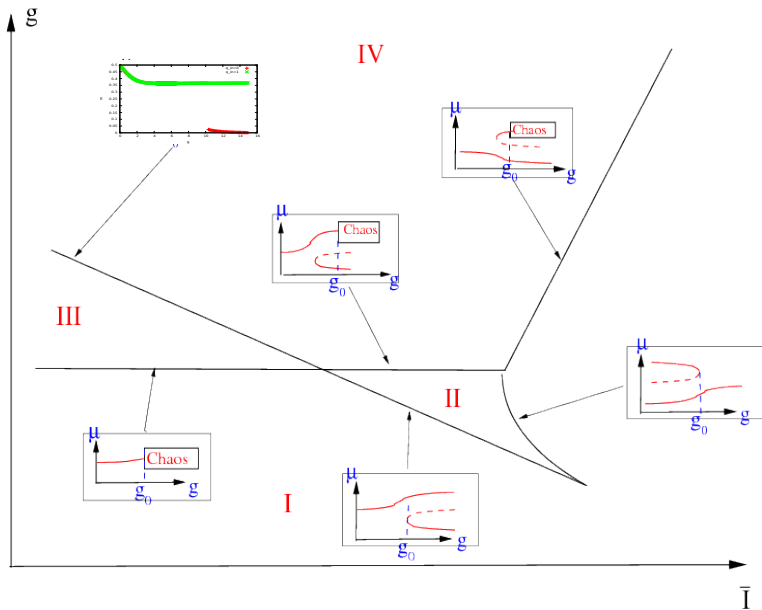
Mean-field analysis and bifurcation map

Cessac et al, 1994; Muratori et al, 2013



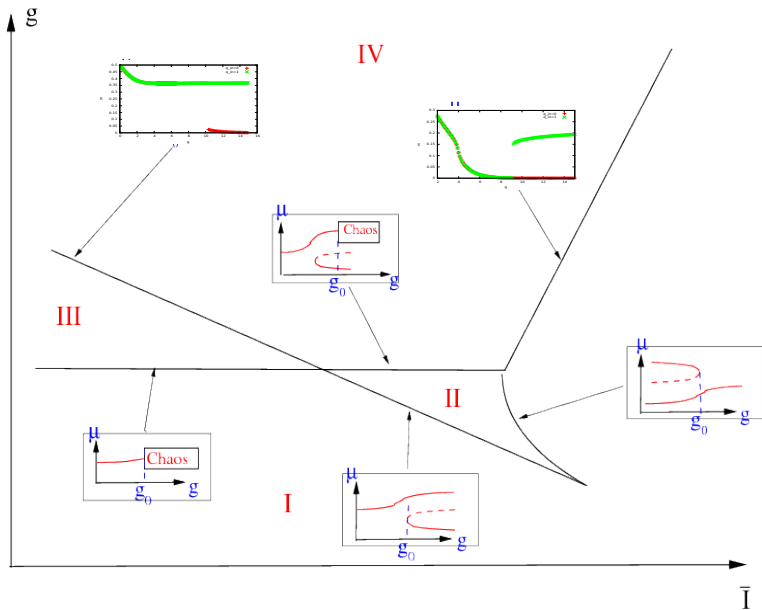
Mean-field analysis and bifurcation map

Cessac et al, 1994; Muratori et al, 2013



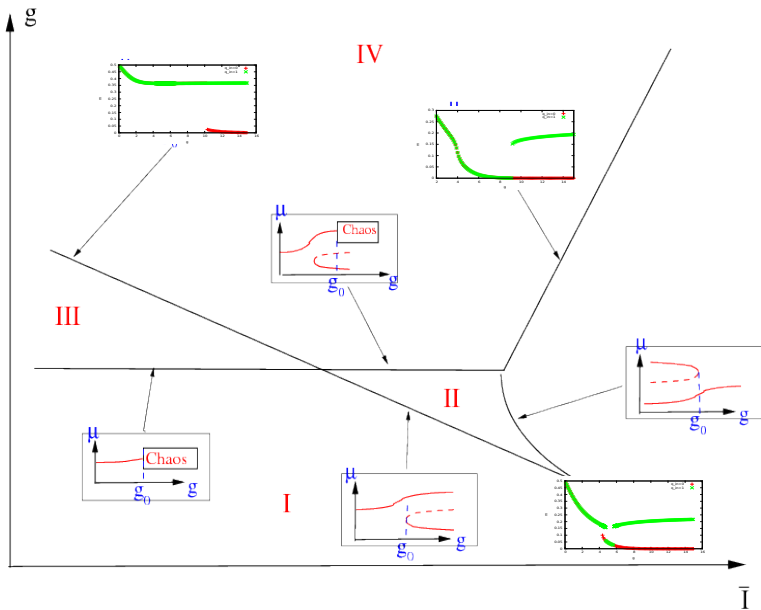
Mean-field analysis and bifurcation map

Cessac et al, 1994; Muratori et al, 2013



Mean-field analysis and bifurcation map

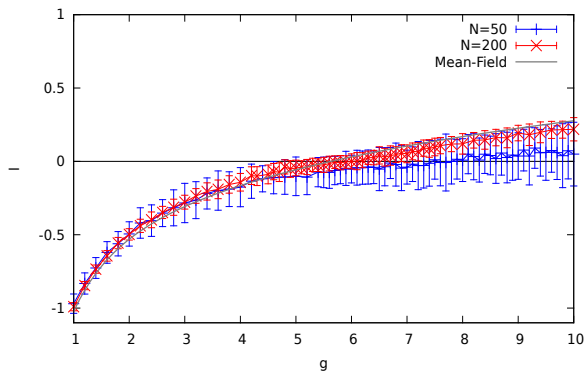
Cessac et al, 1994; Muratori et al, 2013



Maximal Lyapunov exponent

Cessac, 1995; Muratori et al, in preparation

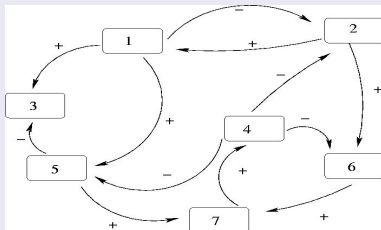
$$\mathbb{E}[\lambda] = \frac{1}{2} \log \left[J^2 \int_{-\infty}^{+\infty} f'_g \left(h\sqrt{v} + \bar{J}\mu - \sigma_\theta^2 \right)^2 Dh \right]$$



Linear response to stimuli and effective connectivity

Linear response to stimuli and effective connectivity

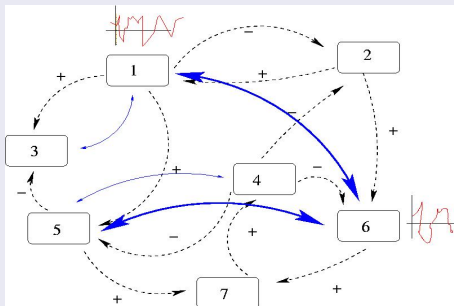
Is there a natural notion of connectivity in the SNT ?



Of course !!! **THE synaptic graph !!!**

Linear response to stimuli and effective connectivity

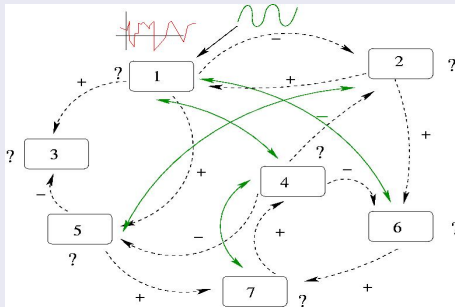
Is there a natural notion of connectivity in the SNT ?



Yes, but this connectivity does not take into account dynamics. What about **correlations** ?

Linear response to stimuli and effective connectivity

Is there a natural notion of connectivity in the SNT ?



Yes, but correlations are non causal. What about **(linear) response** ?

Linear response to stimuli and effective connectivity

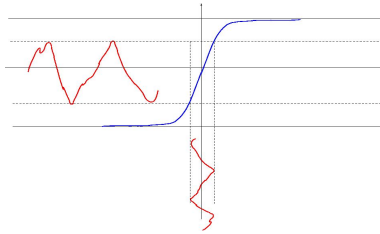


Figure: Amplification

Linear response to stimuli and effective connectivity

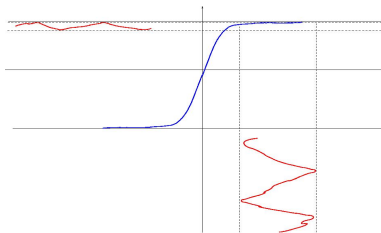


Figure: Saturation

Linear response to stimuli and effective connectivity

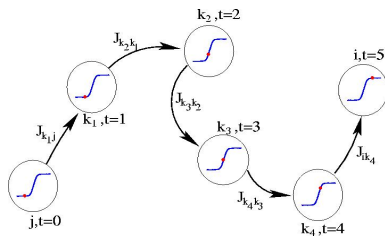


Figure: Propagation

Linear response to stimuli and effective connectivity

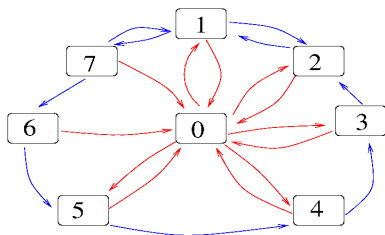


Figure: Effective connectivity ?

Superimpose a weak, time dependent signal $\xi_i(t)$ upon the chaotic background.

$$u_i(t+1) = \gamma u_i(t) + \sum_{j=1}^N J_{ij} f(u_j(t)) - \theta_i + \xi_i(t).$$

What to expect ?

- There will be no measurable effect due to chaos.
- If there is a linear response it obeys the fluctuation-dissipation theorem.

Linear response

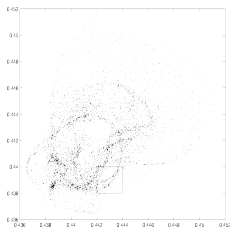


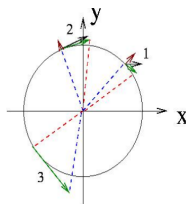
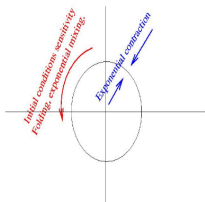
Figure: Strange attractor

Main properties

- Local expansion (unstable manifold);
- Fractal (stable manifold);
- Ergodicity;
- Carries a natural probability measure (Sinai-Ruelle-Bowen) ρ .

Example

$$\begin{cases} r_{t+1} = \gamma r_t + (1 - \gamma) \\ \theta_{t+1} = 2\theta_t \pmod{2\pi} \end{cases}$$



$$DF_{r,\theta}^t = \begin{bmatrix} \gamma^t & 0 \\ 0 & 2^t \end{bmatrix}$$

Figure: Perturbation

Averaging

- The effect of a stimulus can be measured by taking *averages* (time average or ρ average).

Averaging

- The effect of a stimulus can be measured by taking *averages* (time average or ρ average).
- Let ϕ be some function of neurons state. The application of a time dependent stimulus will change the average value of ϕ . If the stimulus is weak, perturbation reads:

$$\delta_t \rho[\phi] = \langle K * \xi(t), \phi \rangle,$$

where the linear response kernel K decomposes as $K = K^s + K^u$.

Averaging

- The effect of a stimulus can be measured by taking *averages* (time average or ρ average).
- Let ϕ be some function of neurons state. The application of a time dependent stimulus will change the average value of ϕ . If the stimulus is weak, perturbation reads:

$$\delta_t \rho[\phi] = \langle K * \xi(t), \phi \rangle,$$

where the linear response kernel K decomposes as $K = K^s + K^u$.

- The unstable part K^u obeys the fluctuation-dissipation theorem; not K^s

Averaging

- The effect of a stimulus can be measured by taking *averages* (time average or ρ average).
- Let ϕ be some function of neurons state. The application of a time dependent stimulus will change the average value of ϕ . If the stimulus is weak, perturbation reads:

$$\delta_t \rho[\phi] = \langle K * \xi(t), \phi \rangle,$$

where the linear response kernel K decomposes as $K = K^s + K^u$.

- The unstable part K^u obeys the fluctuation-dissipation theorem; not K^s
- The Fourier transform of K has poles, corresponding to resonances. The resonances of K^u are in the power spectrum, not those of K^s .

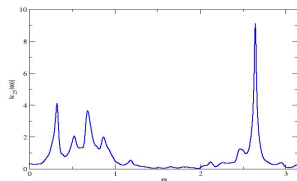


Figure: Power Spectrum

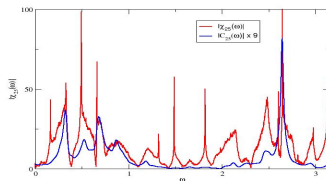
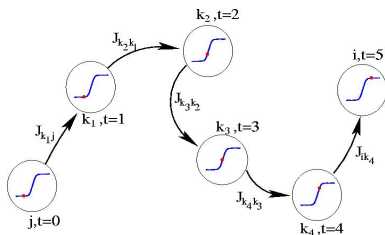


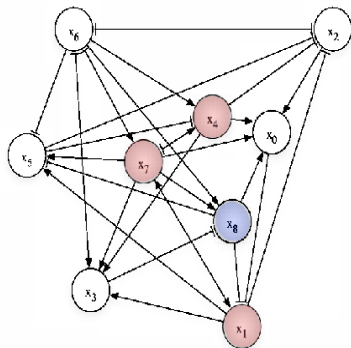
Figure: Susceptibility spectrum

In the SNT model the linear response of neurons membrane potential is given by:

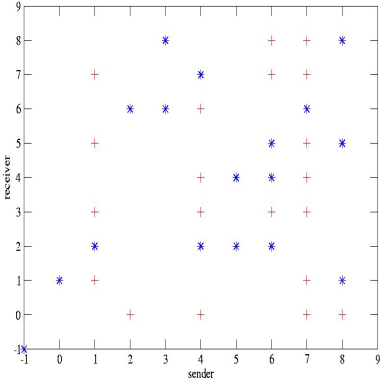
$$K_{ij}(t) = \sum_{\gamma_{ij}(t)} \prod_{l=1}^t J_{k_l k_{l-1}} \rho \left[\prod_{l=1}^t f'(u_{k_{l-1}}(l-1)) \right]$$



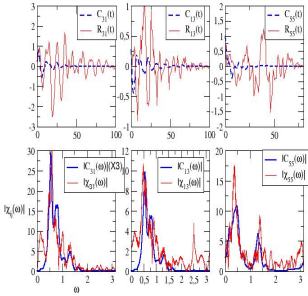
Effective connectivity



Effective connectivity

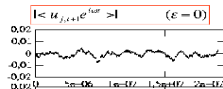
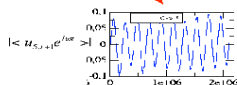
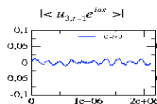
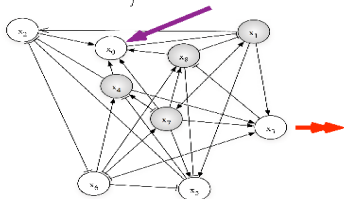


Effective connectivity



Effective connectivity

$$u_{0,t+1} = \sum_j J_{0,j} f(u_{j,t}) + \varepsilon \cos(\omega_M t) \sin(\omega t) \quad (\varepsilon \sim 10^{-3})$$

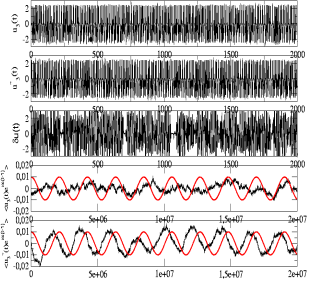


$$\langle u_{S,t+1} \rangle = \varepsilon \chi_{S_0}(\omega) \cos(\omega_M t) \sin(\omega t + \phi_{S_0}(\omega)) + O(\varepsilon^2)$$

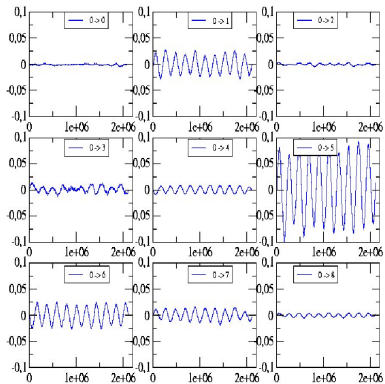
Effective connectivity

$$7 \rightarrow 3, \theta = 0.57$$

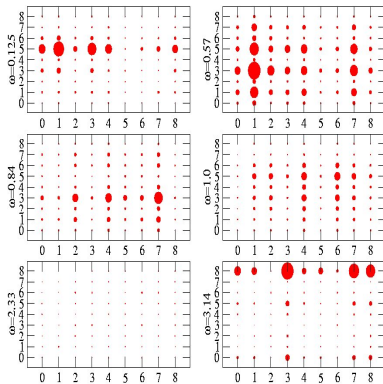
$$\omega_0 = 2.97 \cdot 10^6, \epsilon = 10^{-3}, T = 10^8$$



Effective connectivity

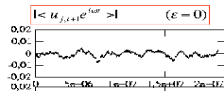
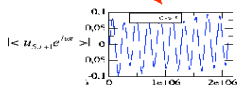
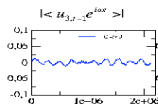
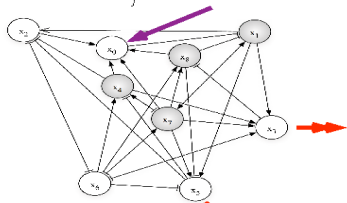


Effective connectivity



Effective connectivity

$$u_{0,t+1} = \sum_j J_{0,j} f(u_{j,t}) + \varepsilon \cos(\omega_M t) \sin(\omega t) \quad (\varepsilon \sim 10^{-3})$$



$$\langle u_{x,t+1} \rangle = \varepsilon \chi_{x0}(\omega) \cos(\omega_M t) \sin(\omega t + \phi_{x0}(\omega)) + O(\varepsilon^2)$$