Lecture 5:
State space analysis

Jochen Braun

Otto-von-Guericke-Universität Magdeburg,
Cognitive Biology Group

Theoretical Neuroscience II, SS 2020
Credits:
H.R. Wilson, "Spikes, Decisions and Actions", 1999

Lecture 5: State space analysis

Today formal analysis of dynamical systems (linear or non-linear, two or more state variables). Linear systems have (at most) one steady-state $\boldsymbol{v}_{ss}$. Time-development $\boldsymbol{v}(t)$ in terms eigenvectors of connectivity matrix $\boldsymbol{A}$. Constant input shifts the steady-state away from 0 (origin). Analyze 2D linear systems in terms of 'nullclines', 'fixed points', and eigenvalues $\lambda_{1,2}$. Possible dynamics are asymptotically stable ('steady-state', $\mathbf{Re}(\lambda_{1,2} < 0)$, neutrally stable ($\mathbf{Re}(\lambda_{1,2} = 0)$, unstable (at least one $\mathbf{Re}\lambda_{1,2} > 0$), spiralling ($\mathbf{Im}(\lambda_1) = \mathbf{Im}(\lambda_1) \neq 0$). Non-linear systems may have multiple steady-states. Analyze in terms of 'nullclines', 'fixed points', and 'Jacobean matrix' (evaluated at each fixed point). Possible dynamics are asymptotically stable (all $\mathbf{Re}(\lambda_i < 0)$, neutrally stable (all $\mathbf{Re}(\lambda_i = 0)$, unstable (at least one $\mathbf{Re}\lambda_i > 0$), spiralling (some $\mathbf{Im}(\lambda_i) \neq 0$). State space analysis is very general and widely useful!

Organization of lecture

- **1. Linear recurrent networks (reminder)**
- **2. Two-dimensional linear systems**
  - **A first example**
  - **Linear feedback in the retina**
- **3. Dynamical possibilities**
- **4. Non-linear dynamics**
- **5. Biological examples**
  - **Divisive gain control**
  - **Short-term memory**

This lecture is a 'fast course' in dynamical systems and contains several fully worked examples, each deserving closer study!

1. Dynamic equations of linear systems

We modify recurrent neural network dynamics (with relaxation)

$$\tau \frac{d\boldsymbol{v}}{dt} = -\boldsymbol{v} + \boldsymbol{M} \cdot \boldsymbol{v} = -\boldsymbol{I} \cdot \boldsymbol{v} + \boldsymbol{M} \cdot \boldsymbol{v} = \underbrace{(-\boldsymbol{I} + \boldsymbol{M})}_{\boldsymbol{A}} \cdot \boldsymbol{v}$$

and use identity matrix $\boldsymbol{I}$ to obtain linear systems dynamics (without relaxation).

Linear system without constant input:

$$\tau \frac{d\boldsymbol{v}}{dt} = \boldsymbol{A} \cdot \boldsymbol{v}, \qquad 0 = \boldsymbol{A} \cdot \boldsymbol{v}_{ss} \quad \Leftrightarrow \quad \boldsymbol{v}_{ss} = 0$$

Linear system with constant input $\boldsymbol{b}$ (inverse matrix $\boldsymbol{A}^{-1}$):

$$\tau \frac{d\boldsymbol{v}}{dt} = \boldsymbol{A} \cdot \boldsymbol{v} + \boldsymbol{b}, \qquad 0 = \boldsymbol{A} \cdot \boldsymbol{v}_{ss} + \boldsymbol{b} \quad \Leftrightarrow \quad \boldsymbol{v}_{ss} = -\boldsymbol{A}^{-1} \cdot \boldsymbol{b}$$

Coordinate shift nulls constant input:

$$\boldsymbol{u} = \boldsymbol{v} - \boldsymbol{v}_{ss} \quad \Rightarrow \quad \tau \frac{d\boldsymbol{u}}{dt} = \boldsymbol{A} \cdot \boldsymbol{u}, \qquad 0 = \boldsymbol{A} \cdot \boldsymbol{u}$$

Linear recurrent neural networks (reminder)

The dynamic equation of a linear recurrent network

$$\tau \frac{d\boldsymbol{v}}{dt} = -\boldsymbol{v} + \boldsymbol{M} \cdot \boldsymbol{v} = (\boldsymbol{M} - \boldsymbol{I}) \cdot \boldsymbol{v},$$

is solved by an eigenvector expansion:

$$\boldsymbol{v}(t) = \sum_{\mu=1}^{N} c_\mu(t)\, \boldsymbol{e}_\mu, \qquad c_\mu(t) = c_\mu(0)\, e^{-(1-\lambda_\mu)t/\tau}$$

If a steady-state exists (i.e., if $\lambda_\mu < 1\ \forall \mu$), the steady-state is

$$\boldsymbol{v}_{ss} = 0$$

In terms of the eigenvalues and eigenvectors of the new matrix $\boldsymbol{A}$

$$\boldsymbol{A} = \boldsymbol{M} - \boldsymbol{I}$$

$$(1 - \lambda)\, \boldsymbol{e} = \boldsymbol{A} \cdot \boldsymbol{e} \qquad \Leftrightarrow \qquad \lambda\, \boldsymbol{e} = \boldsymbol{M} \cdot \boldsymbol{e}$$

the new linear dynamics, with new state vector $\boldsymbol{x}$ replacing $\boldsymbol{v}$, is

$$\tau \frac{d\boldsymbol{x}}{dt} = \boldsymbol{A} \cdot \boldsymbol{x}$$

has the solution

$$\boldsymbol{x}(t) = \sum_{\mu=1}^{N} c_\mu(t)\, \boldsymbol{e}_\mu, \qquad c_\mu(t) = c_\mu(0)\, e^{\lambda_\mu t/\tau}, \qquad \boldsymbol{x}_{ss} = 0$$

(Note that λ_{new} replaces $(1 - \lambda_{old})$.)

Adding a constant input $\boldsymbol{b}$

$$\tau \frac{d\boldsymbol{x}'}{dt} = \boldsymbol{A} \cdot \boldsymbol{x}' + \boldsymbol{b}$$

shifts the steady-state from $\boldsymbol{x}_{ss} = 0$ to

$$\boldsymbol{x}'_{ss} = -\left(\boldsymbol{A}^{-1}\right) \cdot \boldsymbol{b} \qquad \Leftrightarrow \qquad 0 = \boldsymbol{A} \cdot \boldsymbol{x}'_{ss} + \boldsymbol{b}$$

provided that $\boldsymbol{A}^{-1}$ exists.

Solutions with and without constant input $\boldsymbol{b}$ differ by the constant offset $\boldsymbol{x}'_{ss}$:

$$\boldsymbol{x}'(t) = \boldsymbol{x}(t) + \boldsymbol{x}'_{ss} = \sum_{\mu=1}^{N} c_\mu(t)\, \boldsymbol{e}_\mu + \boldsymbol{x}'_{ss}$$

Summary

- Linear recurrent networks have (at most) one steady-state $\boldsymbol{x}_{ss}$.
- The time-evolution $\boldsymbol{x}(t)$ can be obtained in terms of the eigenvectors of the connectivity matrix $\boldsymbol{A} = \boldsymbol{M} - \boldsymbol{I}$.
- The time-dependent coefficients are exponential functions of the eigenvalues.

$$c_\mu(t) = c_\mu(0)\, e^{\lambda_\mu\, t/\tau}$$

- Without constant input, the steady-state is $\boldsymbol{x}_{ss} = 0$ (coordinate origin).
- Constant input $\boldsymbol{b}$ shifts the steady-state to $\boldsymbol{x}'_{ss} = -(\boldsymbol{A}^{-1}) \cdot \boldsymbol{b}$.

2. Two-dimensional linear system

The two-dimensional linear system ($\tau = 1$ for convenience)

$$\underbrace{\dot{\boldsymbol{x}} = \frac{d\boldsymbol{x}}{dt}}_{\textit{new notation}!} = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{b} \qquad \text{or} \qquad \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

has the general solution

$$\boldsymbol{x}(t) = C_1\,\boldsymbol{e}_1\,e^{\lambda_1 t} + C_2\,\boldsymbol{e}_2\,e^{\lambda_2 t} + \boldsymbol{x}_{ss}$$

where $\lambda_{1,2}$ are the *eigenvalues*

and $\boldsymbol{e}_{1,2}$ are the *eigenvectors* of matrix $\boldsymbol{A}$.

$\boldsymbol{x}_{ss}$ is the steady-state and $C_{1,2}$ follow from the initial condition $\boldsymbol{x}_0$.

The eigenvalues $\lambda_{1,2}$ are the (real or complex) solutions of the *characteristic equation*

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = \left| \begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{pmatrix} \right| = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12} a_{21} = 0$$

The eigenvectors (normalized to unit length) are (if $a_{12} \neq 0$)

$$\boldsymbol{e}_1 = \frac{1}{\sqrt{a_{12}^2 + (\lambda_1 - a_{11})^2}} \begin{bmatrix} a_{12} \\ \lambda_1 - a_{11} \end{bmatrix}$$

$$\boldsymbol{e}_2 = \frac{1}{\sqrt{a_{12}^2 + (\lambda_2 - a_{11})^2}} \begin{bmatrix} a_{12} \\ \lambda_2 - a_{11} \end{bmatrix}$$

Matlab function **eig()** performs both calculations.

Nullclines and fixed points

Nullclines are states at which one state variable is stationary:

$$\dot{x} \stackrel{!}{=} 0 \qquad \text{or} \qquad \dot{y} \stackrel{!}{=} 0$$

In our system, the *nullclines* are straight lines with equations

$$\begin{aligned} 0 &\stackrel{!}{=} \dot{x} = a_{11}x + a_{12}y + b_1 \\ 0 &\stackrel{!}{=} \dot{y} = a_{21}x + a_{22}y + b_2 \end{aligned}$$

Fixed points are states at which both state variables are stationary:

$$0 \stackrel{!}{=} \dot{\boldsymbol{x}} = \boldsymbol{A} \cdot \boldsymbol{x} + \boldsymbol{b} \qquad \Leftrightarrow \qquad \boldsymbol{x}_{ss} = -\left(\boldsymbol{A}^{-1}\right) \cdot \boldsymbol{b}$$

In our case, there is at most *one* fixed point (intersection of straight lines). **We wonder if and when a fixed point is also a steady-state?!**

Summary so far

The time-evolution of the state $\boldsymbol{x}(t)$ of a two-dimensional linear system

$$\dot{\boldsymbol{x}} = \boldsymbol{A} \cdot \boldsymbol{x} + \boldsymbol{b}, \qquad x(0) = \boldsymbol{x}_0$$

may be analyzed in term of

- nullclines $0 = a_{11}x + a_{12}y + b_1$ and $0 = a_{21}x + a_{22}y + b_2$ (both straight lines)
- fixed point $\boldsymbol{x}_{ss} = -\left(\boldsymbol{A}^{-1}\right) \cdot \boldsymbol{b}$ (intersection of nullclines)
- eigenvalues λ_1 and λ_2 of $\boldsymbol{A}$
- general solution

$$\boldsymbol{x}(t) = C_1\, \boldsymbol{e}_1\, e^{\lambda_1 t} + C_2\, \boldsymbol{e}_2\, e^{\lambda_2 t} + \boldsymbol{x}_{ss}$$

2.1 A first example

As a concrete example, consider the linear system

$$\dot{\boldsymbol{x}} = \boldsymbol{A} \cdot \boldsymbol{x} + \boldsymbol{b} \qquad \boldsymbol{A} = \begin{pmatrix} -9 & -5 \\ 1 & -3 \end{pmatrix} \qquad \boldsymbol{b} = \begin{pmatrix} 1 \\ 7 \end{pmatrix}$$

The fixed point (obtained by *backslash* in Matlab) is

$$\boldsymbol{x}_{ss} = -\boldsymbol{A}^{-1} \cdot \boldsymbol{b} = -\boldsymbol{A} \backslash b = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

The characteristic equation for the *eigenvalues* $\lambda_{1,2}$ is

$$|\boldsymbol{A} - \lambda\, \boldsymbol{I}| = \left| \begin{pmatrix} -9-\lambda & -5 \\ 1 & -3-\lambda \end{pmatrix} \right| = (-9-\lambda)(-3-\lambda) + 5 = 0$$

with the solutions

$$\lambda_1 = -4, \qquad \lambda_2 = -8$$

To determine the *eigenvector* associated with λ_i, we set $\lambda_i\, \boldsymbol{e}_i = \boldsymbol{A} \cdot \boldsymbol{e}_i$

$$\lambda_i \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} -9 & -5 \\ 1 & -3 \end{pmatrix} \cdot \begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

$$\lambda_i\, x_i = -9\, x_i - 5\, y_i \qquad \Leftrightarrow \qquad \frac{x_i}{y_i} = \frac{-5}{9 + \lambda_i}$$

Normalizing the *eigenvectors* to unit length, we obtain

$$\boldsymbol{e}_i = -\frac{1}{\sqrt{5^2 + (9 + \lambda_i)^2}} \begin{pmatrix} -5 \\ 9 + \lambda_i \end{pmatrix}$$

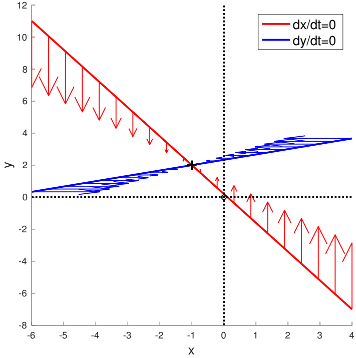
or

$$\boldsymbol{e}_1 = \frac{1}{\sqrt{5^2 + (9 - 4)^2}} \begin{pmatrix} -5 \\ 9 - 4 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$\boldsymbol{e}_2 = \frac{1}{\sqrt{5^2 + (9 - 8)^2}} \begin{pmatrix} -5 \\ 9 - 8 \end{pmatrix} = \frac{1}{\sqrt{26}} \begin{pmatrix} -5 \\ 1 \end{pmatrix}$$

Nullclines and fixed point, in state space

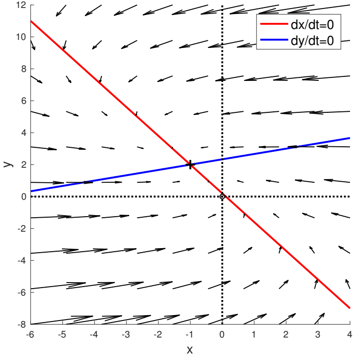
$$\left.\begin{array}{l} 0 \stackrel{!}{=} \dot{x} = -9x - 5y + 1 \quad \Leftrightarrow \quad y = \frac{1}{5}(1 - 9x) \\ 0 \stackrel{!}{=} \dot{y} = x - 3y + 7 \quad \Leftrightarrow \quad y = \frac{1}{3}(x + 7) \end{array}\right\} \quad \boldsymbol{x}_{ss} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$



The x, y-plane is the 'state space' (space of all possible states).

x- and y-gradients, in state space

$$\dot{x} = -9x - 5y + 1, \qquad \dot{y} = x - 3y + 7$$



The x, y-plane is the 'state space' (space of all possible states).
Here we see the direction of development from all possible states.

Coefficients of general solution

From the general solution and the initial condition $\boldsymbol{x}(0)$

$$\boldsymbol{x}(t) = C_1\, \boldsymbol{e}_1\, e^{\lambda_1 t} + C_2\, \boldsymbol{e}_2\, e^{\lambda_2 t} + \boldsymbol{x}_{ss}$$

we obtain the coefficients $C_{1,2}$

$$\boldsymbol{x}(0) = C_1\, \boldsymbol{e}_1 + C_2\, \boldsymbol{e}_2 + \boldsymbol{x}_{ss} = \boldsymbol{E} \cdot \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} + \boldsymbol{x}_{ss}, \qquad \boldsymbol{E} = [\boldsymbol{e}_1 \boldsymbol{e}_2]$$

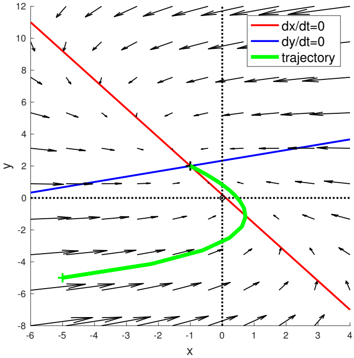
$$\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \boldsymbol{E}^{-1} \cdot [\boldsymbol{x}(0) - \boldsymbol{x}_{ss}]$$

Trajectory

Plotting the general solution

$$\boldsymbol{x}(t) = C_1\, \boldsymbol{e}_1\, e^{\lambda_1 t} + C_2\, \boldsymbol{e}_2\, e^{\lambda_2 t} + \boldsymbol{x}_{ss}$$

for $\boldsymbol{x}(0) = [-5, -5]$



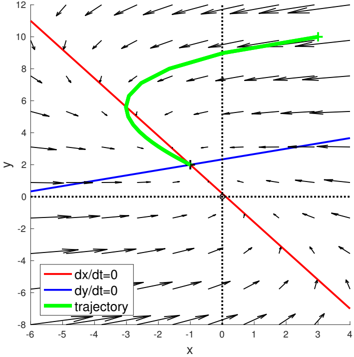
Here we see the 'trajectory' of development from one initial condition (green cross).

Trajectory

Plotting the general solution

$$\boldsymbol{x}(t) = C_1\,\boldsymbol{e}_1\,e^{\lambda_1 t} + C_2\,\boldsymbol{e}_2\,e^{\lambda_2 t} + \boldsymbol{x}_{ss}$$

for $\boldsymbol{x}(0) = [3, 10]$



Here we see the 'trajectory' of development from another initial condition (green cross).

2.2 Another example: negative feedback in the retina

In the retina, cones stimulate horizontal cells, which in turn provide inhibitory feedback to cones. According to Schapf et al. (1990), the dynamic response of cones $C(t)$ and horizontal cells $H(t)$ to a step change in luminance L, is well approximated by a two-dimensional linear system:

$$\frac{dC}{dt} = \dot{C} = \frac{1}{\tau_C}\left(-C - kH + L\right)$$

$$\frac{dH}{dt} = \dot{H} = \frac{1}{\tau_H}\left(-H + C\right)$$

$$\Leftrightarrow \quad \dot{\boldsymbol{R}} = \boldsymbol{A} \cdot \boldsymbol{R} + \boldsymbol{B}, \qquad \boldsymbol{R} \equiv \begin{bmatrix} C \\ H \end{bmatrix}, \qquad \boldsymbol{A} \equiv \begin{bmatrix} -1/\tau_C & -k/\tau_C \\ 1/\tau_H & -1/\tau_H \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} L/\tau_C \\ 0 \end{bmatrix}$$

with $\tau_C = 0.025s$, $\tau_H = 0.08s$, and $k = 4$.

The fixed point is easy to find: $\dot{H} = 0 \quad \Rightarrow \quad H_{ss} = C_{ss}$

$\dot{C} = 0 \quad \Rightarrow \quad C_{ss} = L/(1+k)$

With Matlab function **eig()**, we find the *eigenvalues* and *eigenvectors* to be complex:

$$\lambda_1 = -26.25 + 42.56\, i, \qquad \lambda_2 = -26.25 - 42.56\, i$$

$$\boldsymbol{E}_1 = \begin{bmatrix} -0.9631 \\ 0.0828 + 0.2563\, i \end{bmatrix}, \qquad \boldsymbol{E}_2 = \begin{bmatrix} -0.9631 \\ 0.0828 - 0.2563\, i \end{bmatrix}$$

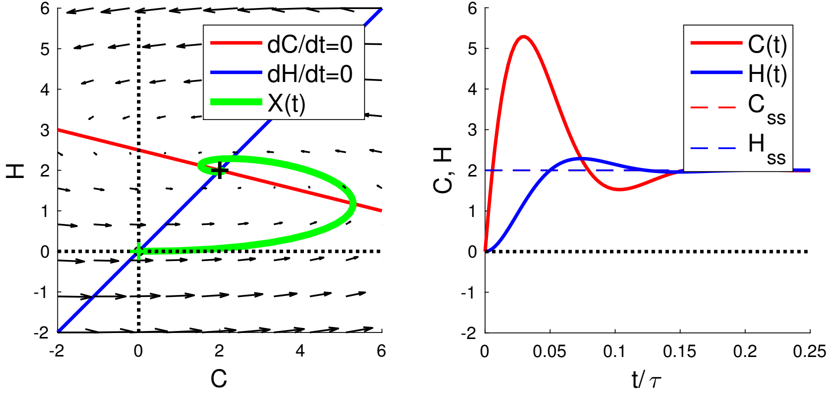
When the initial condition $\boldsymbol{R}_0$ is real, the coefficients $C_{1,2}$ are such that the imaginary parts cancel:

$$\boldsymbol{R}(t) = Re\left[C_1\, \boldsymbol{E}_1\, \cos(\lambda_1 t) + C_2\, \boldsymbol{E}_2\, \cos(\lambda_2 t)\right] + \boldsymbol{R}_{ss}$$

$$0 = Im\left[C_1\, \boldsymbol{E}_1\, i\, \sin(\lambda_1 t)\right] + Im\left[C_2\, \boldsymbol{E}_2\, i\, \sin(\lambda_2 t)\right]$$

where we have used Euler's formula $e^{i\,\phi} = \cos\phi + i\,\sin\phi$.

Luminance L steps from 0 to 10



Left: state space trajectory. Right: Time-courses. Both show dampened oscillation converging on fixed point!

3. Dynamical possibilities

We can now see that all possible two-dimensional linear systems fall into a small group of categories.

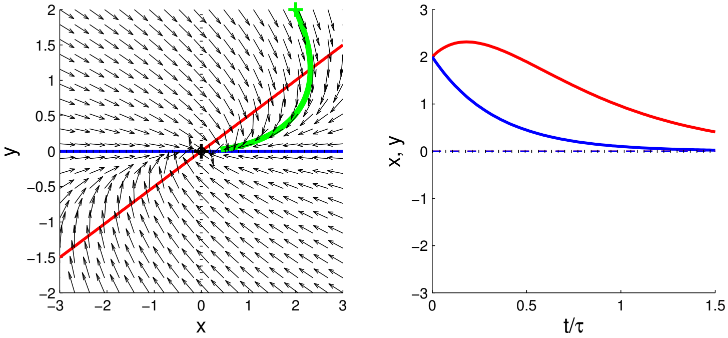
There is at most one stable fixed point (intersection of the nullclines). As the characteristic equation of the recurrent connectivity matrix $\boldsymbol{A}$ is quadratic, it has exactly two roots, real or complex, which may be equal. These roots are the eigenvalues and determine the eigenvectors.

The dynamical behavior around the fixed point depends on the eigenvalues:

- Asymptotically stable: $Re(\lambda_1)$ and $Re(\lambda_2)$ are negative.
- Neutrally stable: $Re(\lambda_1)$ and $Re(\lambda_2)$ are zero.
- Unstable: At least one of $Re(\lambda_1)$ and $Re(\lambda_2)$ is positive.
- Spiralling: $Im(\lambda_1) = -Im(\lambda_2)$ are nonzero.

Asymptotically stable (real parts negative):

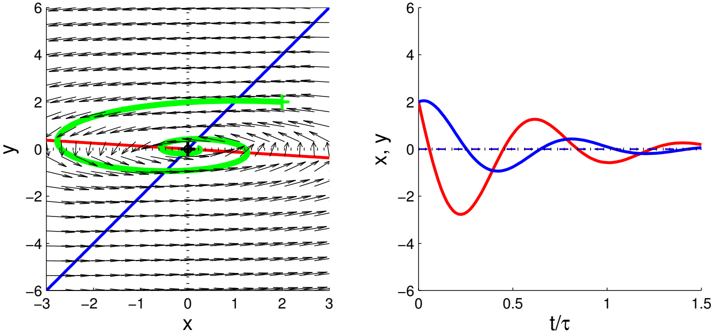
$$\dot{\mathbf{x}} = \begin{pmatrix} -2 & 4 \\ 0 & -3 \end{pmatrix} \mathbf{x}, \qquad \lambda_{1,2} = -2, -3$$



Trajectories decay exponentially to fixed point and **steady-state**.

Asymptotically stable with spiraling (real parts negative):

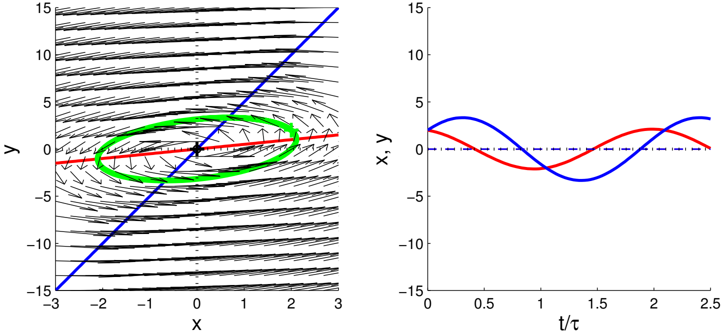
$$\dot{\boldsymbol{x}} = \begin{pmatrix} -2 & -16 \\ 4 & 2 \end{pmatrix} \boldsymbol{x}, \qquad \lambda_{1,2} = -2 \pm 8\, i$$



Trajectories spiral to the fixed point and **steady-state**.

Neutrally stable with spiralling (real parts zero):

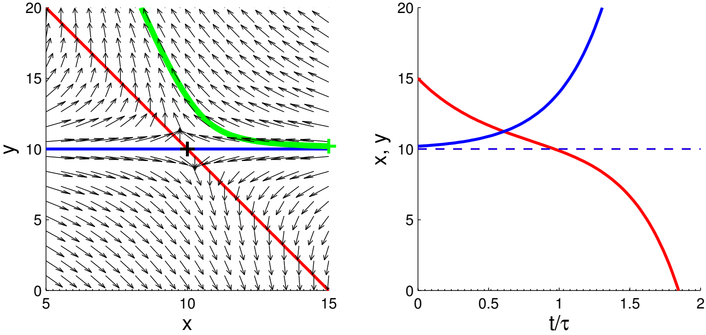
$$\dot{\boldsymbol{x}} = \begin{pmatrix} 1 & -2 \\ 5 & -1 \end{pmatrix} \boldsymbol{x}, \qquad \lambda_{1,2} = \pm 3\, i$$



Trajectories orbit the fixed point. There is no steady-state!

Unstable (at least one real part positive):

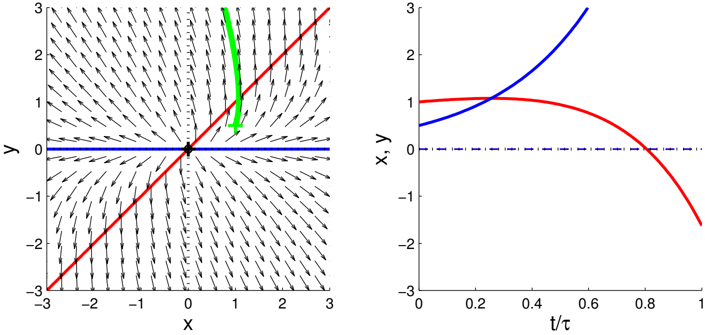
$$\dot{\boldsymbol{x}} = \begin{pmatrix} -2 & -1 \\ 0 & 3 \end{pmatrix} \boldsymbol{x}, \qquad \lambda_{1,2} = -2, 3$$



Trajectories grow exponentially away from fixed point. No SS!

Unstable (both real parts positive):

$$\dot{\boldsymbol{x}} = \begin{pmatrix} 1 & -1 \\ 0 & 3 \end{pmatrix} \boldsymbol{x}, \qquad \lambda_{1,2} = 1, 3$$



Trajectories grow exponentially away from fixed point. No SS!

Higher-dimensional linear systems

These results generalize to N-dimensional linear systems

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{b} \quad \text{or} \quad \tau \begin{bmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_N \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a1N \\ \vdots & \ddots & \vdots \\ a_{N1} & \dots & a_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}$$

has the general solution

$$\boldsymbol{x}(t) = \sum_{i=1}^{N} C_i\,\boldsymbol{e}_i\,e^{\lambda_i t} + \boldsymbol{x}_{ss}$$

where λ_i are the *eigenvalues* and $\boldsymbol{e}_i$ are the *eigenvectors* of matrix $\boldsymbol{A}$.

$\boldsymbol{x}_{ss}$ is the fixed point (and potential steady-state) and the constants C_i are set by initial conditions $\boldsymbol{x}_0$.

Summary linear dynamical systems

There is at most one stable fixed point:

$$\boldsymbol{x}_{ss} = -\boldsymbol{A}^{-1} \cdot \boldsymbol{b}$$

The characteristic equation of the recurrent connectivity matrix $\boldsymbol{A}$ has N roots, real or complex. These roots are the eigenvalues and determine the eigenvectors.

The dynamical behavior around the fixed point depends on the eigenvalues.

- Asymptotically stable 'steady-state': All $Re(\lambda_i)$ are negative.
- Neutrally stable: All $Re(\lambda_i)$ are zero.
- Unstable: At least one of $Re(\lambda_i)$ is positive.
- Spiralling: Some $Im(\lambda_i)$ are nonzero.

4. Non-linear dynamics

The insights of the previous section regarding linear recurrent networks can also help us to understand the non-linear recurrent networks in the brain.

Specifically, we will see that the qualitative behavior in the vicinity of ‘fixed points’ remains unchanged.

Real parts of eigenvalues determine nature of ‘fixed points’ (steady-states or otherwise).

The main novelty will be the appearance of multiple ‘fixed points’, sometimes with very different behavior.

Dynamic equations

A general description of two interacting neural populations is

$$\frac{dx}{dt} = F(x, y)$$

$$\frac{dy}{dt} = G(x, y)$$

where $x(t)$ and $y(t)$ are the activations (state variables) and $F(x, y)$ and $G(x, y)$ are non-linear functions of the activations, where we have absorbed characteristic times and constant inputs into $F()$ and $G()$.

To be physiologically plausible, $F()$ and $G()$ will be finite and continuous (or have a finite number of discontinuities).

Nullclines

The two 'nullclines' of the system are the sets of points defined by

$$F(x, y) = 0$$

or

$$G(x, y) = 0$$

In general, each equation defines a curve (or line) in state space.

On each curve, one of the activations (state variables) is stationary (time-derivative is zero).

Fixed points

The fixed points, which are potential steady-states, are the intersections of the two isoclines, i.e., points for which

$$F(x, y) = 0 \qquad \text{and} \qquad G(x, y) = 0$$

As the isoclines may be curved, there may be several such intersections.

Isoclines and fixed points are essential tools for analyzing also nonlinear dynamics.

Stability of fixed point

Consider a two-dimensional non-linear system with fixed point $\boldsymbol{X}_{ss}$

$$\dot{\boldsymbol{X}} = \boldsymbol{F}(\boldsymbol{X}) \qquad \boldsymbol{F}(\boldsymbol{X}_{ss}) = 0$$

In the vicinity of $\boldsymbol{X}_{ss}$, we can define a *linear approximation*

$$\dot{\boldsymbol{X}} = \boldsymbol{A}\,\boldsymbol{X} \qquad \boldsymbol{A} \equiv \begin{pmatrix} \frac{\delta F}{\delta x} & \frac{\delta F}{\delta y} \\ \frac{\delta G}{\delta x} & \frac{\delta G}{\delta y} \end{pmatrix}$$

where A is the *Jacobian matrix* (or simply *Jacobian*) and all partial derivatives are evaluated *at the fixed point*!

Near the fixed point, the linear approximation will behave like the original non-linear system. Accordingly, linear and non-linear fixed point will exhibit the same kind of stability (stable, neutral, unstable, spiralling, etc).

Higher-dimensional systems

The same logic applies to a fixed point $\boldsymbol{X}_{ss}$ in a higher-dimensional system

$$\dot{\boldsymbol{X}} = \boldsymbol{F}(\boldsymbol{X}) \qquad\qquad \boldsymbol{F}(\boldsymbol{X}_{ss}) = 0$$

In the vicinity of $\boldsymbol{X}_{ss}$, we can construct a *linear approximation*

$$\dot{\boldsymbol{X}} = A\,\boldsymbol{X} \qquad\qquad \boldsymbol{A} \equiv \left.\begin{pmatrix} \frac{\delta F_1}{\delta x_1} & \frac{\delta F_1}{\delta x_2} & \cdots \\ \frac{\delta F_2}{\delta x_1} & \frac{\delta F_2}{\delta x_2} & \cdots \\ \vdots & \vdots & \frac{\delta F_N}{\delta x_N} \end{pmatrix}\right|_{\boldsymbol{X}=\boldsymbol{X}_{ss}}$$

by evaluating the partial derivatives in the *Jacobian matrix* at the fixed point.

Types of stability

Therefore, the stability of the fixed points of non-linear systems falls into the familiar categories, revealed by the eigenvalues:

- Asymptotically stable: All $Re(\lambda_i)$ are negative.
- Neutrally stable: All $Re(\lambda_i)$ are zero.
- Unstable: At least one of $Re(\lambda_i)$ is positive.
- Spiralling: Some $Im(\lambda_i)$ are nonzero.

Summary

- The most important tool for analyzing non-linear recurrent systems are *nullclines*.
- These are the subspaces (curves for 2D systems) in which one of the state variables is stationary.
- The intersection of *nullclines* identifies *fixed points*, where all state variables are stationary.
- In the vicinity of each fixed point, the system behavior can be analyzed by linearization.
- A linear system with the *Jacobian matrix* (evaluated at a fixed point) captures the behavior of the non-linear system.

5. Biological examples

As a first example, we consider a non-linear gain control network with *divisive* inhibition (as opposed to *subtractive* inhibition). Divisive gain control is a common model for the interaction between nearby cortical columns.

As a second example, we consider a short-term memory network, the activity of which retains a memory of a recent episode of transient stimulation.

Divisive gain control

We consider one excitatory and one inhibitory population, with activities $E(t)$ and $I(t)$ and constant external input S:

$$\frac{dE}{dt} = \frac{1}{\tau_E} \left(-E + \frac{S}{1+I} \right), \qquad E(t) \geq 0$$

$$\frac{dI}{dt} = \frac{1}{\tau_I} \left(-I + 2\,E \right), \qquad I(t) \geq 0$$

A stimulus S linearly excites population E, which in turn linearly excites population I. Population I provides a non-linear, inhibitory feedback: it divisively diminishes stimulation S.

This is termed *divisive inhibition* or *divisive gain control*.

Fixed point

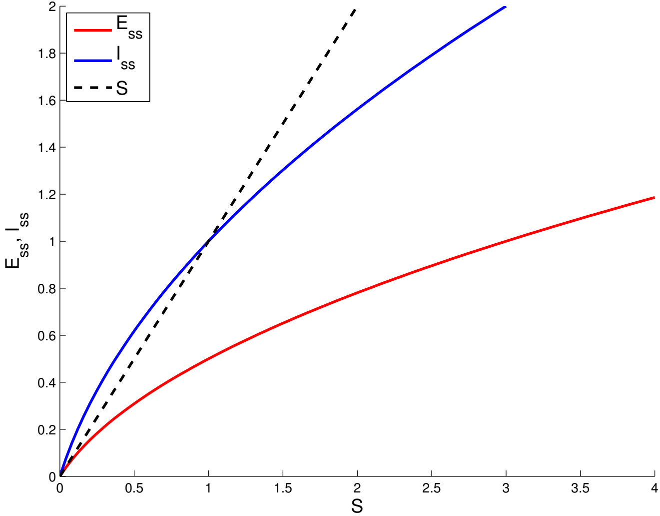
Both $E(t)$ and $I(t)$ relax to fixed point (steady-state) values

$$E_{ss} = \frac{S}{1 + I_{ss}}, \qquad I_{ss} = 2\,E_{ss}$$

so that

$$\Rightarrow \qquad E_{ss} + 2E_{ss}^2 = S, \qquad E_{ss} = -\frac{1}{4} + \sqrt{\frac{S}{2} + \frac{1}{16}}$$

Note that activities remain positive ($E(t), I(t) \geq 0$), provided that $E(0), I(0), S \geq 0$.



With growing stimulus S, steady-state activity E_{ss} at first grows proportionately and later less than proportionately. Steady-state activity I_{ss} is always larger than E_{ss}.

Nullclines

To begin the dynamical analysis, we obtain the nullclines:

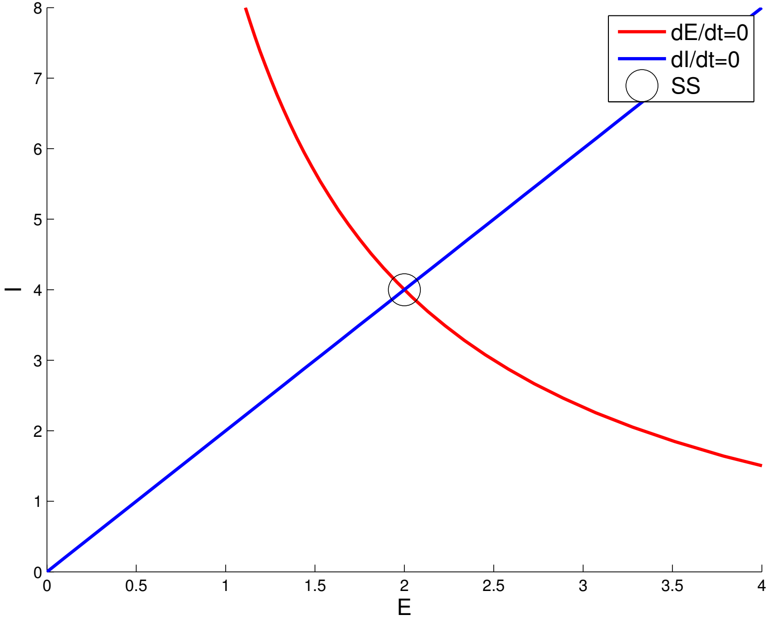
$$0 \stackrel{!}{=} \frac{dE}{dt} = \frac{1}{\tau_E}\left(-E + \frac{S}{1+I}\right) \qquad \Rightarrow \qquad I = \frac{S}{E} - 1$$

$$0 \stackrel{!}{=} \frac{dI}{dt} = \frac{1}{\tau_I}\left(-I + 2\,E\right) \qquad \Rightarrow \qquad I = 2\,E$$

The steady-state (see above) is the intersection:

$$E_{ss} + 2\,E_{ss}^2 = S \qquad \Rightarrow \qquad E_{ss} = \frac{-1+\sqrt{8\,S+1}}{4}, \quad I_{ss} = 2E_{ss}$$

$S = 10$



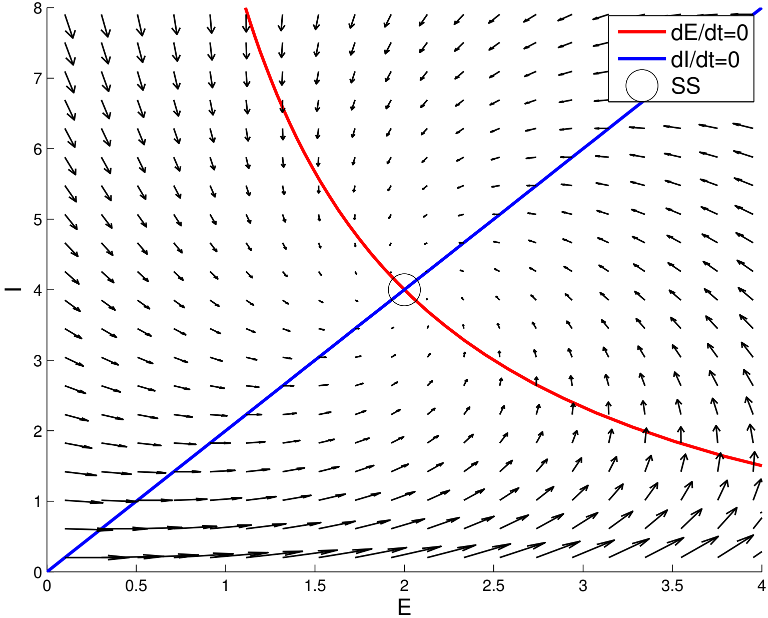
Dynamical trajectories

As a next step, we compute the gradient vectors $(\dot{E}, \dot{I})$ for each state vector (E, I), to obtain the direction of trajectories:

$$\begin{pmatrix} \dot{E} \\ \dot{I} \end{pmatrix} = \begin{pmatrix} F_E \\ F_I \end{pmatrix},$$

$$F_E \equiv \frac{1}{\tau_E} \left(-E + \frac{S}{1+I} \right), \qquad F_I \equiv \frac{1}{\tau_I} \left(-I + 2\,E \right)$$

Gradient vectors (F_E, F_I)



Analyze fixed point

The trajectory gradients already indicate that the fixed point is stable. To confirm the characteristics of the fixed point, we obtain the Jacobian matrix

$$A = \left. \begin{pmatrix} \frac{\delta F_E}{\delta E} & \frac{\delta F_E}{\delta I} \\ \\ \frac{\delta F_I}{\delta E} & \frac{\delta F_I}{\delta I} \end{pmatrix} \right|_{(E_{ss}, I_{ss})}$$

$$F_E \equiv \frac{1}{\tau_E} \left(-E + \frac{S}{1+I} \right), \qquad F_I \equiv \frac{1}{\tau_I} \left(-I + 2\,E \right)$$

$$\frac{\delta F_E}{\delta E} = -\frac{1}{\tau_E}, \qquad \frac{\delta F_E}{\delta I} = -\frac{1}{\tau_E} \, \frac{S}{(1+I_{ss})^2}$$

$$\frac{\delta F_I}{\delta E} = \frac{2}{\tau_I}, \qquad \frac{\delta F_I}{\delta I} = -\frac{1}{\tau_I}$$

For example, assuming $S = 10$ and $\tau_E = \tau_I = 10ms$, we obtain

$$A = \begin{pmatrix} -\frac{1}{10} & -\frac{1}{(1+I_{ss})^2} \\ \\ \frac{1}{5} & -\frac{1}{10} \end{pmatrix}$$

With

$$E_{ss} = \frac{-1 + \sqrt{8\,S + 1}}{4} = \frac{-1 + 9}{4} = 2, \quad I_{ss} = 2E_{ss} = 4$$

this becomes

$$A = \begin{pmatrix} -\frac{1}{10} & -\frac{1}{25} \\ \\ \frac{1}{5} & -\frac{1}{10} \end{pmatrix} \qquad \text{with} \qquad \lambda_{1,2} = -0.1 \pm 0.089\,\mathrm{i}$$

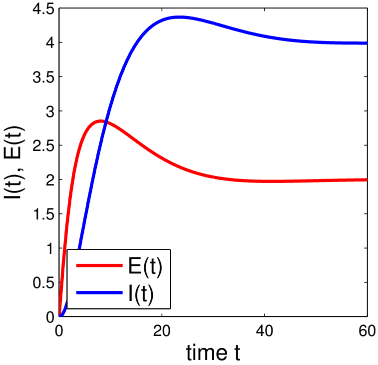
so that with have an *asymptotically stable spiral point*.

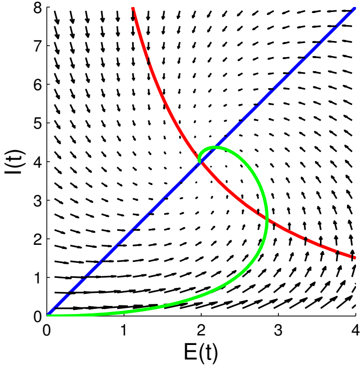
Onset responses

The network responds to a stimulus onset with a brief excitatory overshoot before settling to its new steady-state.

The size and shape of this overshoot depends on the characteristic times of the excitatory and inhibitory populations.

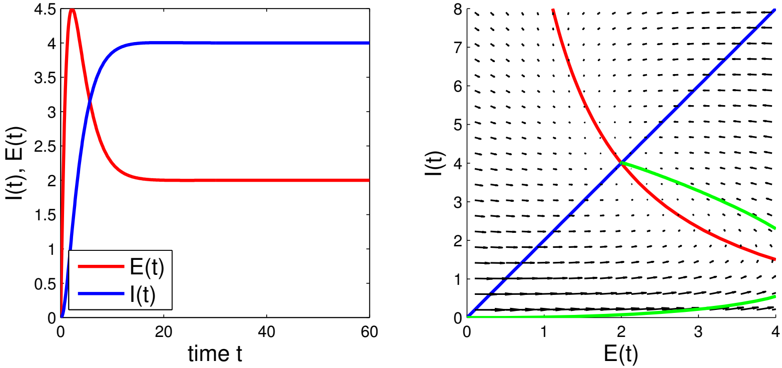
Onset response $\tau_E = \tau_I = 10$





The response to stimulus onset is dampened quickly by inhibition.

Onset response $\tau_E = 2$, $\tau_I = 10$



Reducing the time-constant of excitation makes the onset response more pronounced.

Summary

- Divisive inhibition helps to map a large input range (e.g., luminance) onto a smaller output range (e.g., neuronal firing).
- The steady-state output is proportional to a small power (e.g., square root) of the input.
- After input onset, the system trajectory describes an excursion before settling on the new steady-state.
- The size and duration of this excursion depends on the characteristic times of excitatory and inhibitory populations.

Short-term memory

Visual neurons in inferotemporal cortex respond selectively when particular objects are flashed (Fuster, 1995). What interests us here is that they *continue to respond* even after the object has disappeared again. The response stops only when another object is flashed.

Thus, the activity of these neurons seems to maintain a *short-term memory* of the most recent object. We use a non-linear recurrent network to model this property.

We use two populations, $E_{1,2}$ that excite each other. The dynamic equations are

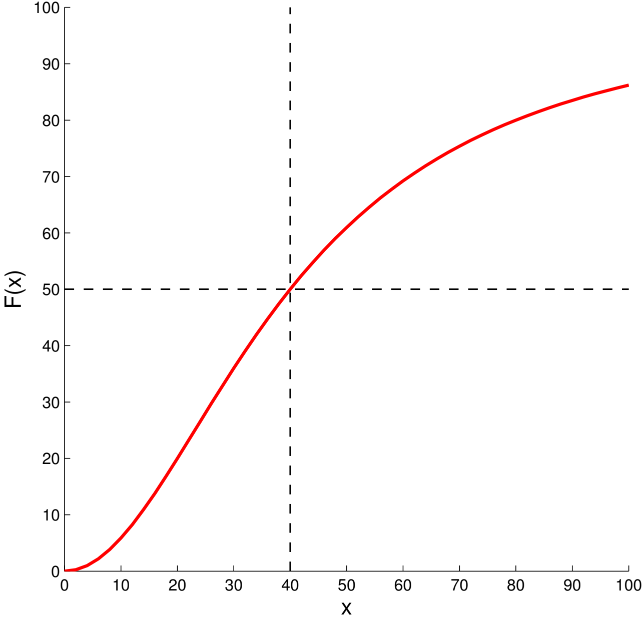
$$\frac{dE_1}{dt} = \frac{1}{\tau}\left[-E_1 + F(E_2)\right]$$
$$\frac{dE_2}{dt} = \frac{1}{\tau}\left[-E_2 + F(E_1)\right]$$

where τ is a characteristic time and $F(x)$ is a non-linear activation function:

$$F(x) = R_{max}\,\frac{x^2}{\kappa^2 + x^2}$$

We choose $R_{max} = 100Hz$, and $\kappa = 40Hz$.

Activation function



Fixed points

Because of symmetry, any fixed points E_{ss} must be on the diagonal $E_1 = E_2$. Specifically, they must satisfy the cubic equation

$$E_{ss} = R_{max} \frac{E_{ss}^2}{\kappa^2 + E_{ss}^2} \qquad \Leftrightarrow \qquad E_{ss}^3 - R_{max}\, E_{ss}^2 + \kappa^2\, E_{ss} = 0$$

$$\Leftrightarrow \quad E_{ss}\, \left[E_{ss} - E'_{ss}\right]\, \left[E_{ss} - E''_{ss}\right] = 0$$

Evidently, the roots of this equation are

$$E_{ss} = 0, \qquad E'_{ss} = \frac{R_{max} + \sqrt{R_{max}^2 - 4\kappa^2}}{2}, \qquad E''_{ss} = \frac{R_{max} - \sqrt{R_{max}^2 - 4\kappa^2}}{2}$$

With $R_{max} = 100Hz$ and $\kappa = 40Hz$, we get

$$E_{ss} = 0Hz \qquad E'_{ss} = 50 + 30 = 80Hz \qquad E''_{ss} = 50 - 30 = 20Hz$$

Isoclines

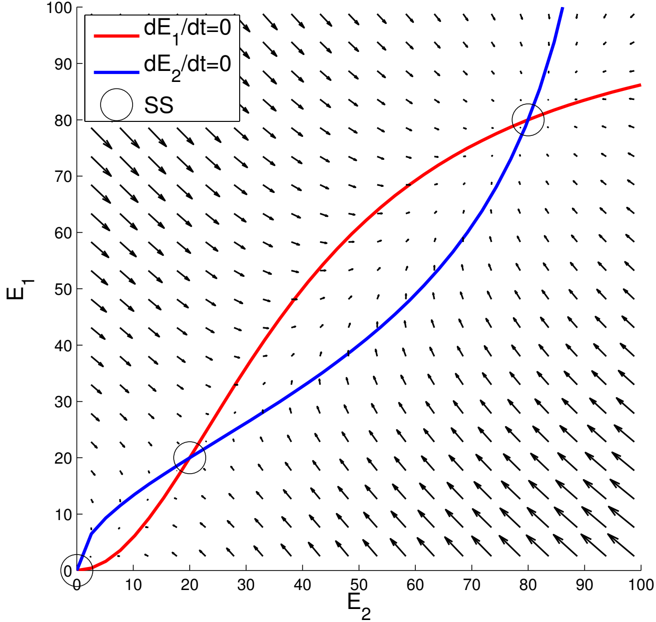
The isoclines are

$$\frac{dE_1}{dt} = 0 \qquad \Rightarrow \qquad E_1 = R_{max} \, \frac{E_2^2}{\kappa^2 \, + E_2^2}$$

$$\frac{dE_2}{dt} = 0 \qquad \Rightarrow \qquad E_2 = R_{max} \, \frac{E_1^2}{\kappa^2 \, + E_1^2}$$

$$\Leftrightarrow \qquad E_2 \, \kappa^2 + E_2 \, E_1^2 = R_{max} \, E_1^2$$

$$E_1 = \kappa \sqrt{\frac{E_2}{R_{max} - E_2}}$$



dE_1/dt=0
dE_2/dt=0
SS
E_1
E_2
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100

Characteristics of fixed points

We have found three fixed points, $E_{ss} = 0Hz$, $20Hz$, and $80Hz$.

To understand the characteristics of each fixed point, we need to compute the *Jacobian matrix* for each point.

The Jacobian describes the limiting behavior of the non-linear system in terms of a linear dynamical system (in the vicinity of the fixed point).

The analysis confirms what the trajectory gradients had already suggested: $0Hz$ and $80Hz$ are stable fixed points, but $20Hz$ is an unstable fixed point.

Jacobian matrices

To determine the nature of our three fixed points, we need to know the eigenvalues of each Jacobian matrix. To this end, we'll need the derivative of our activation function

$$F(x) = R_{max} \, \frac{x^2}{\kappa^2 + x^2}, \qquad \frac{dF(x)}{dx} = R_{max} \, \frac{2\,\kappa^2\,x}{(\kappa^2 + x^2)^2}$$

In general, the system

$$F_1 \equiv \frac{1}{\tau}\left[-E_1 + R_{max}\,\frac{E_2^2}{\kappa^2 + E_2^2}\right]$$

$$F_2 \equiv \frac{1}{\tau}\left[-E_2 + R_{max}\,\frac{E_1^2}{\kappa^2 + E_1^2}\right]$$

has the Jacobian

$$A = \begin{pmatrix} \frac{\delta F_1}{\delta E_1} & \frac{\delta F_1}{\delta E_2} \\ \frac{\delta F_2}{\delta E_2} & \frac{\delta F_2}{\delta E_2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\tau} & \frac{2 R_{max}\,\kappa^2 E_2}{\tau\left(\kappa^2 + E_2^2\right)^2} \\ \frac{2 R_{max}\,\kappa^2 E_1}{\tau\left(\kappa^2 + E_1^2\right)^2} & -\frac{1}{\tau} \end{pmatrix}$$

Jacobian matrices

With $\tau = 20$, $R_{max} = 100$, $\kappa = 40$, this evaluates to

$$A = \begin{pmatrix} -\frac{1}{\tau} & \frac{2R_{max}\kappa^2 E_2}{\tau\left(\kappa^2+E_2^2\right)^2} \\ \frac{2R_{max}\kappa^2 E_1}{\tau\left(\kappa^2+E_1^2\right)^2} & -\frac{1}{\tau} \end{pmatrix} =$$

$$= \begin{pmatrix} -\frac{1}{20} & \frac{16000\, E_2}{\left(1600+E_2^2\right)^2} \\ \frac{16000\, E_1}{\left(1600+E_1^2\right)^2} & -\frac{1}{20} \end{pmatrix}$$

$E_{ss} = 0 Hz$

At the first fixed point, $E_1 = E_2 = E_{ss} = 0$, we obtain

$$A = \begin{pmatrix} -\frac{1}{20} & 0 \\ 0 & -\frac{1}{20} \end{pmatrix}$$

with eigenvalues

$$\lambda_{1,2} = -0.05$$

Thus, the fixed point point (0, 0) is *asymptotically stable*.

$E'_{ss} = 20Hz$

At the second fixed point, $E_{ss} = 20Hz$, we obtain

$$A = \begin{pmatrix} -\frac{1}{20} & \frac{320000}{2000^2} \\ \frac{320000}{2000^2} & -\frac{1}{20} \end{pmatrix} = \begin{pmatrix} -\frac{1}{20} & \frac{2}{25} \\ \frac{2}{25} & -\frac{1}{20} \end{pmatrix}$$

with eigenvalues

$$\lambda_1 = +0.03, \quad \lambda_2 = -0.13$$

Thus, the fixed point point (20, 20) is an *unstable saddle point*.

$E_{ss} = 80Hz$

At the third fixed point, $E_{ss} = 80Hz$, we obtain

$$A = \begin{pmatrix} -\frac{1}{20} & \frac{1280000}{10000^2} \\ \frac{1280000}{10000^2} & -\frac{1}{20} \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{20} & \frac{8}{625} \\ \frac{8}{625} & -\frac{1}{20} \end{pmatrix}$$

with eigenvalues

$$\lambda_1 = -0.07, \quad \lambda_2 = -0.03$$

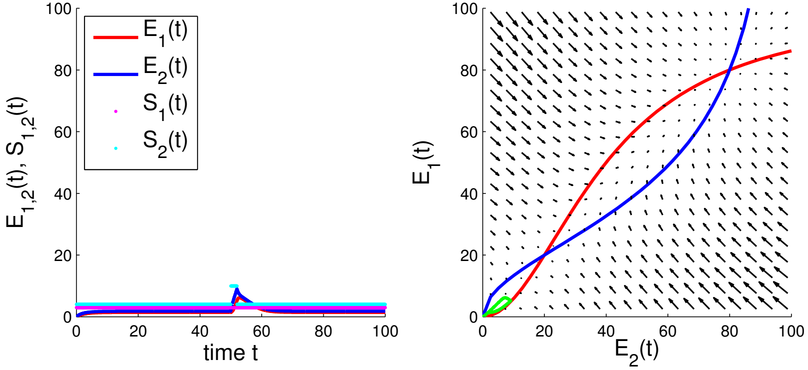
Thus, the fixed point point (80, 80) is again *asymptotically stable*.

Trajectory examples

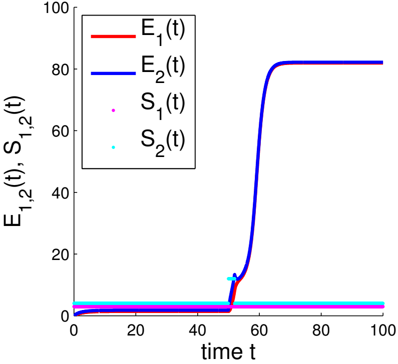
To see how these fixed points influence trajectories, we subject the system to time-varying stimulation $S_{1,2}(t)$. The modified dynamic equations are

$$\frac{dE_1}{dt} = \frac{1}{\tau}\left[-E_1 + F(E_2 + S_1)\right]$$
$$\frac{dE_2}{dt} = \frac{1}{\tau}\left[-E_2 + F(E_1 + S_2)\right]$$

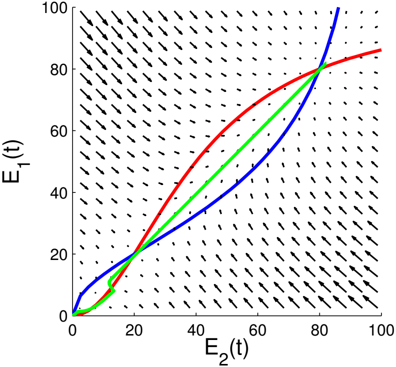
For simplicity, we keep input $S_1(t)$ constant and add to $S_2(t)$ pulses of different amplitude.



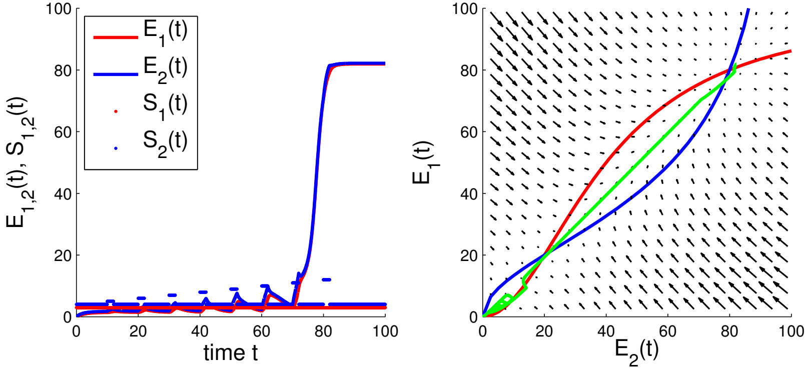
E_1(t)
E_2(t)
S_1(t)
S_2(t)
E_{1,2}(t), S_{1,2}(t)
time t
E_1(t)
E_2(t)



100
80
60
40
20
0
0 20 40 60 80 100
time t
$E_{1,2}(t)$, $S_{1,2}(t)$
$E_1(t)$
$E_2(t)$
$S_1(t)$
$S_2(t)$



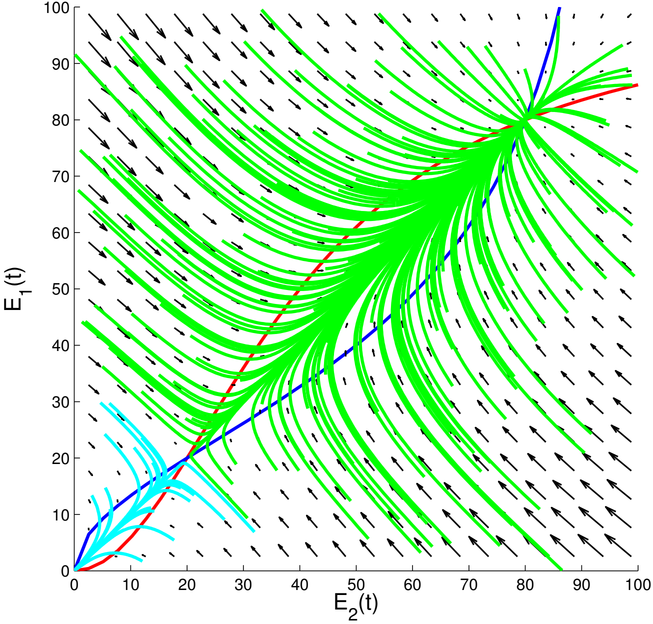
100
80
60
40
20
0
0 20 40 60 80 100
$E_2(t)$
$E_1(t)$


100
80
60
40
20
0
0
20
40
60
80
100
time t
$E_{1,2}(t), S_{1,2}(t)$
$E_1(t)$
$E_2(t)$
$S_1(t)$
$S_2(t)$
$E_1(t)$
$E_2(t)$

Basins of attraction

Starting trajectories at random points of the state space and following them to one two stable fixed points, we discover that each of these points is surrounded by a *basin of attraction*.

The curve separating these two basins is like a water-shed and is called the *separatrix*.



100
90
80
70
60
50
40
30
20
10
0
E_1(t)
0
10
20
30
40
50
60
70
80
90
100
E_2(t)

Summary

- Visual neurons in inferotemporal cortex are activated by transient stimulation (with particular visual objects) and remain active even after the stimulus has been removed.
- This short-term memory can be modeled by mutual excitation between two populations.
- If mutual excitation is sufficiently strong and the activation function is sufficiently steep, such a network exhibits two stable fixed points.
- The high-activity state maintains itself even after stimulation has been removed.
- This example illustrates some of the dynamical possibilities of non-linear recurrent networks.

Next:
Hebbian plasticity