

# Dynamics of Neural Systems

## Linear Dynamical Systems

Martin A. Giese

[Martin.giese@uni-tuebingen.de](mailto:Martin.giese@uni-tuebingen.de)

Nov 25, 2024

# Overview

- Basic definitions
- Linear dynamical systems
- Stability of linear systems

# Overview

- Basic definitions
- Linear dynamical systems
- Stability of linear systems

# Dynamical systems in neuroscience

- The behavior of individual neurons is characterized by dynamical equations (differential equations (DEQs) describing functions of time)  $\Leftarrow$  HH model.
- Many phenomena at the level of individual neurons, and neural networks can be understood better with methods from dynamical systems theory.
- Many of you might know something about **linear dynamical systems** (DS); unfortunately, most neural phenomena are described by nonlinear DS.
- Nonlinear dynamical systems often also discussed under the keyword 'chaos theory'.

# Dynamical systems (DS): definition

- Dynamical systems describe the behavior of a **state variable**  $\mathbf{x}$  over time.
- Distinction:

**Discrete dynamical systems (map):**  $t$  is an integer.

$$\mathbf{x}[t+1] = \mathbf{f}(\mathbf{x}[t], t)$$

(Iteration of a function; not treated extensively here.)

**Continuous dynamical systems:**  $t \in \mathbb{R}$  is continuous.

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), t) \quad (\text{differential equation in time})$$

- The function  $\mathbf{x}(t)$  (or  $\mathbf{x}[t]$ ) is called **trajectory** or **orbit** of the DS.

# Autonomous dynamical systems

- Dynamical systems are called **autonomous** if the describing equation does not explicitly depend on time:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t)) \quad (\text{respectively } \mathbf{x}[t+1] = \mathbf{f}(\mathbf{x}[t]))$$

- By introduction of a new dynamic variable non-autonomous dynamical systems can always be made autonomous, e.g.:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), t) \quad \Leftrightarrow \quad \begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{f}(\mathbf{x}(t), y) \\ \frac{dy}{dt} &= 1 \quad y(0) = 0 \end{aligned}$$

Non-autonomous DS

Auxiliary autonomous DS

# Overview

- Basic definitions
- **Linear dynamical systems**
- Stability of linear systems

# Linear dynamical systems I

- General form (for continuous dynamics):

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t)$$

Matrix

Driving term

- We analyze first the **homogenous equation**:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t)$$

- To compute the solution it is handy to define exponentials of matrices (linear operators).



# Exponential of a matrix

- The exponential of a matrix can be defined as:

$$e^{\mathbf{A}} = \sum_{n=0}^{\infty} \frac{1}{n!} \mathbf{A}^n$$

It can be shown that this series converges absolutely for any matrix  $\mathbf{A}$ .

- Like for the scalar case, we have:

$$e^{(a+b)\mathbf{A}} = e^{a\mathbf{A}} e^{b\mathbf{A}}$$

$$e^{\mathbf{A}} e^{-\mathbf{A}} = \mathbf{I}$$

$$e^{-\mathbf{A}} = (e^{\mathbf{A}})^{-1}$$

$$e^{\mathbf{A}} e^{\mathbf{B}} = e^{\mathbf{A}+\mathbf{B}} \quad \text{if } \mathbf{AB} = \mathbf{BA}$$

- Remark: For non-singular  $\mathbf{Q}$

$$e^{\mathbf{Q}\mathbf{A}\mathbf{Q}^{-1}} = \mathbf{Q} e^{\mathbf{A}} \mathbf{Q}^{-1}$$

# Exponential of a matrix

- Remark that for a diagonal matrix  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$  because of  $\Lambda^k = \text{diag}(\lambda_1^k, \dots, \lambda_N^k)$  it follows:

$$e^\Lambda = \text{diag}(\exp(\lambda_1), \dots, \exp(\lambda_N))$$

# Linear dynamical systems II

- From the definition of the series follows:

$$\frac{d}{dt} e^{t\mathbf{A}} = \lim_{h \rightarrow 0} \frac{e^{(t+h)\mathbf{A}} - e^{t\mathbf{A}}}{h} = e^{t\mathbf{A}} \lim_{h \rightarrow 0} \frac{e^{h\mathbf{A}} - \mathbf{I}}{h} = e^{t\mathbf{A}} \mathbf{A} = \mathbf{A} e^{t\mathbf{A}}$$

- This implies that the initial value problem

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) \quad \mathbf{x}(0) = \mathbf{x}_0$$

has the solution:  $\mathbf{x}(t) = e^{t\mathbf{A}} \mathbf{x}_0 = \boldsymbol{\varphi}(t, \mathbf{x}_0)$

,Flow' of the DEQ

(It can be shown that no other solution exists.)

- Interpretation:  $\boldsymbol{\varphi}(t, \mathbf{x}_0)$  defines a **flow field** that maps the initial condition  $\mathbf{x}_0$  onto  $\mathbf{x}(t)$ .

# Linear dynamical systems III

- The solution for the inhomogeneous equation can be found by the technique of **variation of parameters**:
- Ansatz for the solution of the inhomogeneous system:

$$\mathbf{x}(t) = e^{t\mathbf{A}}\mathbf{k}(t)$$

Determine  $\mathbf{k}(t)$  as to match the requirements from inhomogeneous equation:

$$\frac{d\mathbf{x}}{dt} = \underline{\mathbf{A}e^{t\mathbf{A}}\mathbf{k}(t)} + e^{t\mathbf{A}} \frac{d\mathbf{k}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{u}(t) = \underline{\mathbf{A}e^{t\mathbf{A}}\mathbf{k}(t)} + \mathbf{u}(t)$$

$$\frac{d\mathbf{k}}{dt} = e^{-t\mathbf{A}}\mathbf{u}(t) \Rightarrow \mathbf{k}(t) - \mathbf{k}_0 = \int_0^t e^{-t'\mathbf{A}}\mathbf{u}(t') dt'$$

- This implies (because of  $\mathbf{k}_0 = \mathbf{x}_0$ ) for the solution of the inhomogeneous equation:

$$\mathbf{x}(t) = e^{t\mathbf{A}} \left( \mathbf{x}_0 + \int_0^t e^{-t'\mathbf{A}}\mathbf{u}(t') dt' \right) = \underbrace{e^{t\mathbf{A}}\mathbf{x}_0}_{\text{General solution of homog. eq.}} + \underbrace{\int_0^t e^{(t-t')\mathbf{A}}\mathbf{u}(t') dt'}_{\text{Particular solution}}$$

General solution of homog. eq. Particular solution

# Linear dynamical systems IV

- For the special case of a constant input  $\mathbf{u}$  follows (for invertible  $\mathbf{A}$ ):

$$\mathbf{x}(t) = e^{t\mathbf{A}} \left( \mathbf{x}_0 + \int_0^t e^{-t'\mathbf{A}} dt' \mathbf{u} \right) = e^{t\mathbf{A}} \left( \mathbf{x}_0 + \left[ -e^{-t'\mathbf{A}} \mathbf{A}^{-1} \right]_0^t \mathbf{u} \right) \Rightarrow$$

$$\mathbf{x}(t) = e^{t\mathbf{A}} \mathbf{x}_0 - (\mathbf{I} - e^{t\mathbf{A}}) \mathbf{A}^{-1} \mathbf{u} = e^{t\mathbf{A}} (\mathbf{x}_0 + \mathbf{A}^{-1} \mathbf{u}) - \mathbf{A}^{-1} \mathbf{u}$$

- We will see later that for stable systems and any  $\mathbf{b}$

$$\lim_{t \rightarrow \infty} e^{t\mathbf{A}} \mathbf{b} = \mathbf{0}. \text{ This implies: } \mathbf{x}(\infty) = -\mathbf{A}^{-1} \mathbf{u}$$

(The same follows for  $\frac{d\mathbf{x}}{dt} = \mathbf{0}$  from the DEQ.)

# Change of coordinates I

- Assuming that the matrix  $\mathbf{A}$  can be diagonalized (for example if it has only distinct real **eigenvalues**  $\lambda_i$ ), this implies:  $\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{-1}$  where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$ .
- Remark: The columns of  $\mathbf{Q}$  are the eigenvectors of  $\mathbf{A}$ .
- We introduce new coordinates:  $\mathbf{y} = \mathbf{Q}^{-1} \mathbf{x}$ ; in these new coordinates the homogenous DEQ has the form:

$$\frac{d\mathbf{y}}{dt} = \mathbf{\Lambda} \mathbf{y}(t) \quad \Leftrightarrow \quad \frac{dy_i}{dt} = \lambda_i y_i(t)$$

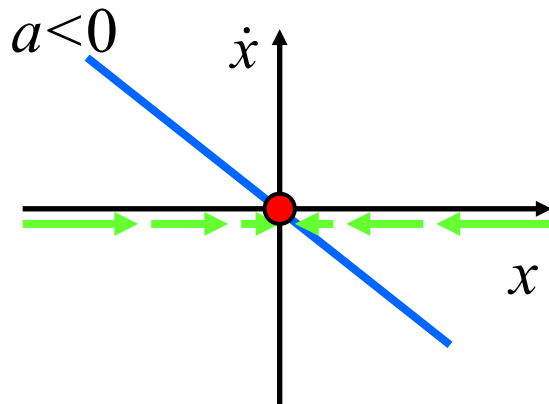
- This is a decoupled set of DEQ, where the individual equations can be easily solved:  $y_i(t) = \exp(\lambda_i t) y_i(0)$ .
- This implies:  $\mathbf{x}(t) = \mathbf{Q} \mathbf{y}(t) = \mathbf{Q} e^{t\mathbf{\Lambda}} \mathbf{y}(0) = \mathbf{Q} e^{t\mathbf{\Lambda}} \mathbf{Q}^{-1} \mathbf{x}(0)$

# Change of coordinates II

- The last point has important consequences:
  1. The solution is a linear combination of the basis functions  $\exp(\lambda_i t)$ .
  2. Each of these 'basis functions' is associated with its own linear subspace, which is given by the columns of  $\mathbf{Q}$ ; the parts of the solutions with different dynamics are thus evolving in separate linear subspaces.
  3. Since  $\mathbf{Q}$  can be shown to have full rank these subspaces are linearly independent.

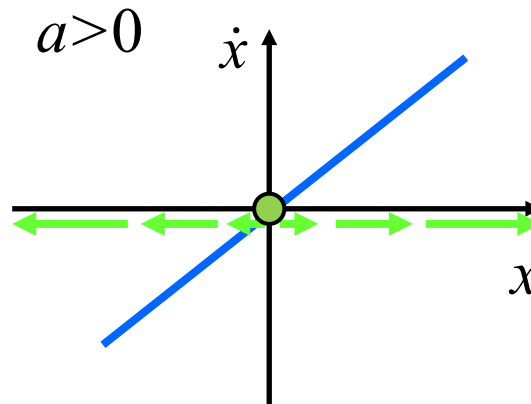
# Phase portrait: 1D systems

- Simple graphical method to analyze how a DS behaves.
- First order equation:  $\frac{dx}{dt} = ax(t)$
- Only three possible situations for **fixed points** with  $\frac{dx}{dt} = 0$ :



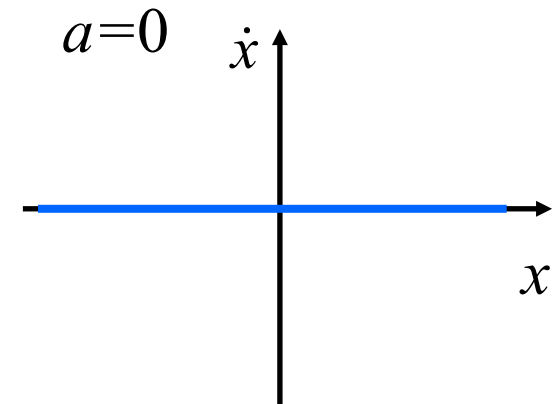
**Stable fixed point**  
/ **attractor for  $x=0$**

- Approached from any initial condition



**Unstable fixed point**  
/ **repellor for  $x=0$**

- Divergence from any initial condition, except for  $x=0$



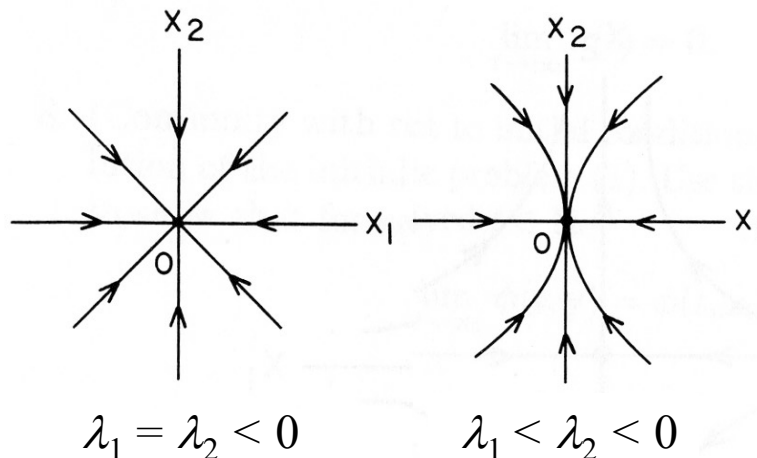
**Marginally stable**

- Initial state persists; but no convergence

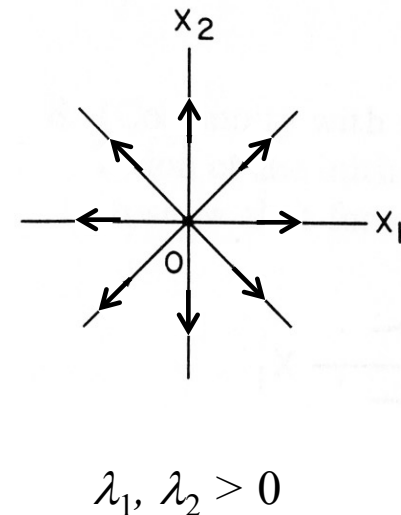


# Phase portrait: 2D systems I

- More interesting / complex behavior for two dimensions.
- Transformed system has two distinct eigenvalues.
- Possible situations for real eigenvalues:

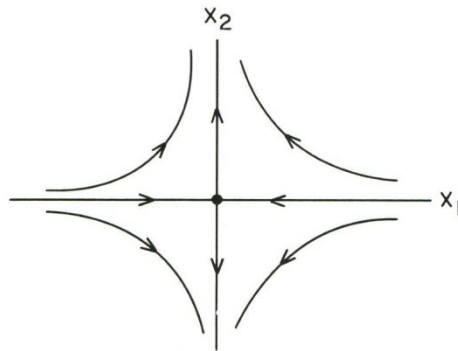


**Stable knodes at  $x=0$**



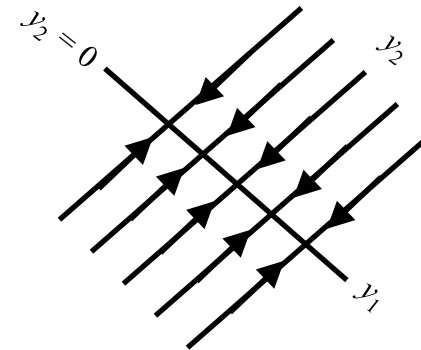
**Unstable knode at  $x=0$**

# Phase portrait: 2D systems II



$$\lambda_1 < 0 < \lambda_2$$

**Saddle point at  $\mathbf{x}=\mathbf{0}$**



$$\lambda_1 = 0, \lambda_2 < 0$$

**Marginally stable manifold  $y_2 = 0$**

- One stable and one unstable direction.
- Stable behavior in the direction of  $y_2$ .
- Marginally stable behavior in the direction of  $y_1$ .
- Behavior results from the superposition of the behavior in different direction in linear independent subspaces.

# Oscillatory components I

- The matrix  $\mathbf{A} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$  has the eigenvalues  $\lambda_{1,2} = a \pm ib$ .

- It can be shown:  $e^{\mathbf{A}t} = e^a \begin{bmatrix} \cos b & -\sin b \\ \sin b & \cos b \end{bmatrix}$

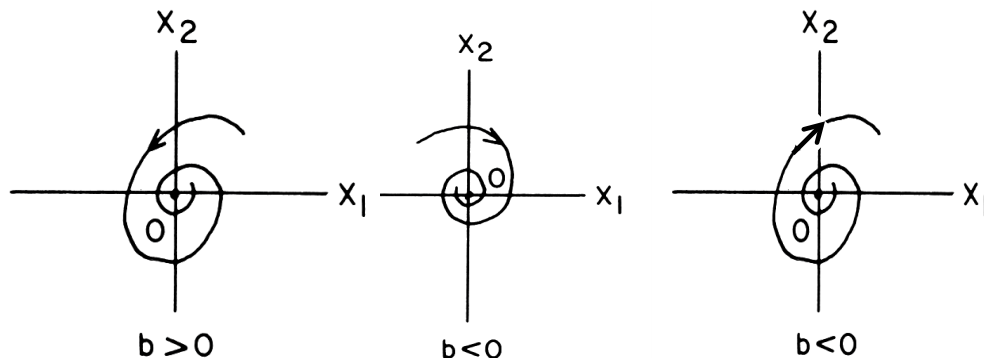
(rotation by angle  $b$  and a stretching by the factor  $e^a$ ).

- This implies for the solution of the corresponding DEQ:

$$\mathbf{x}(t) = e^{\mathbf{A}t} \mathbf{x}_0 = e^{at} \begin{bmatrix} \cos bt & -\sin bt \\ \sin bt & \cos bt \end{bmatrix} \mathbf{x}_0$$

Scaling factor

Rotation matrix



$$\lambda_{1,2} = a \pm ib, a < 0$$

**Stable spirals**

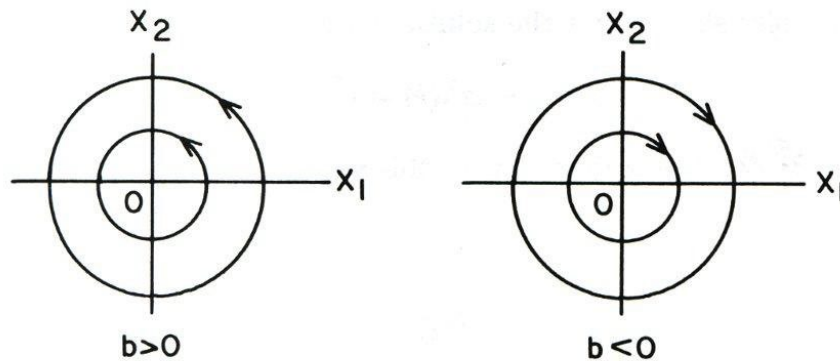
$$\lambda_{1,2} = a \pm ib, a > 0$$

**Unstable spiral**

- These are **spirals** towards or away from the origin.

# Oscillatory components II

- Special case:  $a = 0$ : imaginary eigenvalues; marginally stable oscillation with constant amplitude (**center** at the origin).



$$\lambda_{1,2} = a \pm ib, a = 0$$

**Center at origin**

# Oscillatory components III

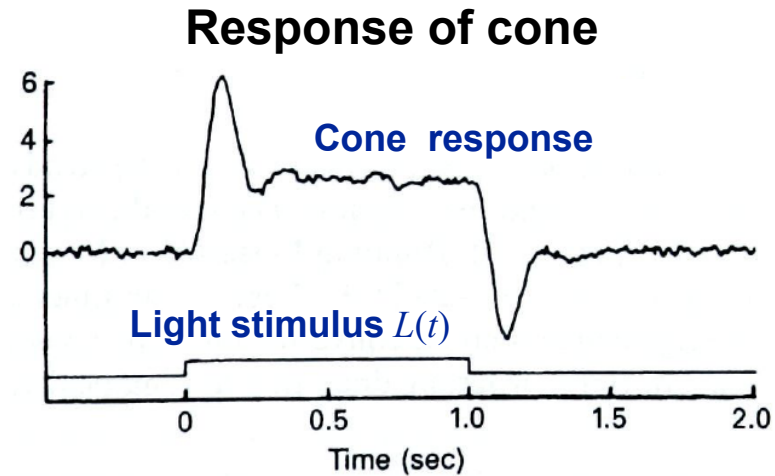
- General result: Matrices with  $N$  distinct complex eigenvalues (appearing in conjugate pairs) can be **(block-)diagonalized** in the form:  $\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^{-1}$  with

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{\Lambda}_N \end{bmatrix} \text{ and } \mathbf{\Lambda}_k = \begin{bmatrix} a_k & -b_k \\ b_k & a_k \end{bmatrix} \quad \text{with} \quad \begin{aligned} \lambda_{k,1} &= a_k + ib_k \\ \lambda_{k,2} &= a_k - ib_k \end{aligned}$$

- The matrix  $\mathbf{Q}$  consists of the real and imaginary parts of the complex eigenvectors of  $\mathbf{A}$ .
- This implies oscillatory solutions in the corresponding linear subspaces if  $b_i \neq 0$ .

# Example: negative feedback in retina I

- Horizontal cells provide negative feedback signals for cones in the retina.
- Recording results for stimulation with rectangular light pulse.
- Oscillatory overshoot.
- Simple model (Wilson book):



**Cone activity:**

**Horizontal cell activity:**

$$\frac{dx_1}{dt} = \frac{1}{\tau_C} (-x_1(t) - kx_2(t) + L(t))$$

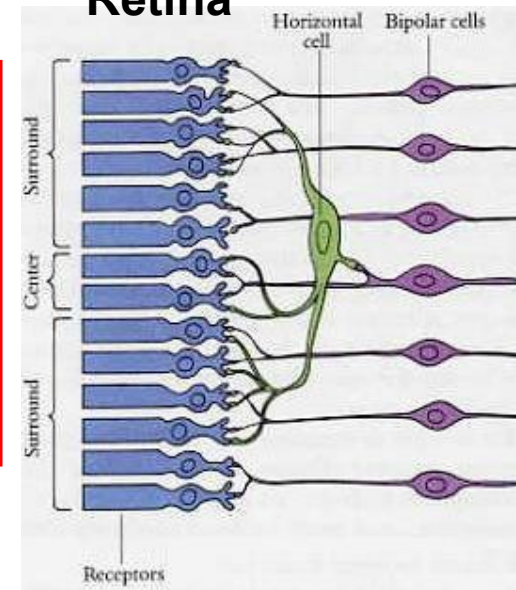
$$\frac{dx_2}{dt} = \frac{1}{\tau_H} (-x_2(t) + x_1(t))$$

with  $\tau_C = 25\text{ms}$ ,  $\tau_H = 80\text{ms}$ , and  $k = 4$

**Light intensity**

**Strength of inhibition**

**Retina**



# Example: negative feedback in retina II

- The system matrix is given by:

$$\mathbf{A} = \begin{bmatrix} -40 & -160 \\ 12.5 & -12.5 \end{bmatrix} \quad (\text{dropping the units})$$

- The eigenvalues are:

$$\lambda_{1,2} = -26.25 \pm 42.56i$$

- This defines a stable spiral.
- Stationary solution for step input

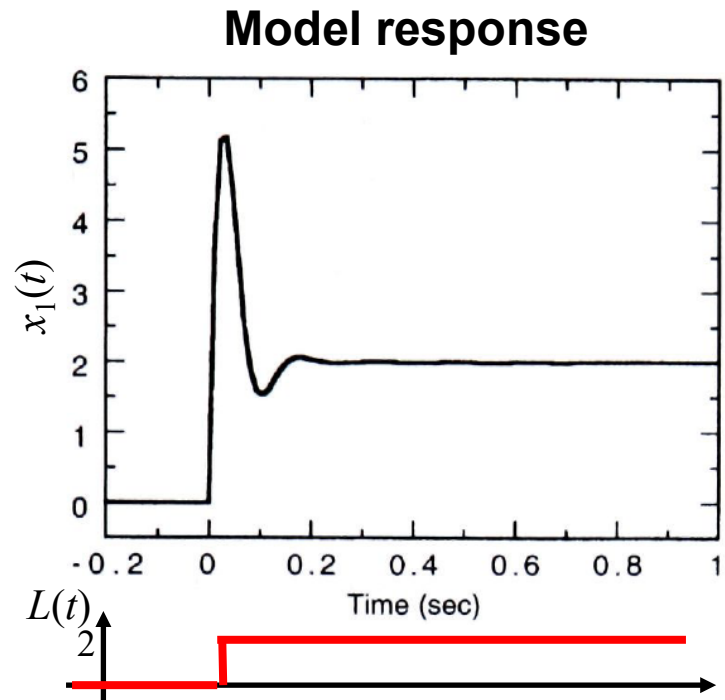
with  $L(t) = 10$  for  $t > 0$  from  $d/dt = 0$ :

$$x_1(\infty) = x_2(\infty) = 2$$

- Explicit solution for non-autonomous system:

$$x_1(t) = \left( -2e^{-t/25\text{ms}} \cos(42.56 \text{ Hz } t) + 8.17e^{-t/25\text{ms}} \sin(42.56 \text{ Hz } t) + 2 \right) \cdot 1(t)$$

$$x_2(t) = \left( -2e^{-t/25\text{ms}} \cos(42.56 \text{ Hz } t) - 1.23e^{-t/25\text{ms}} \sin(42.56 \text{ Hz } t) + 2 \right) \cdot 1(t)$$



# Overview

- Basic definitions
- Linear dynamical systems
- Stability of linear systems



# Dynamical stability

- The linear dynamical system is called **asymptotically stable** if its solution converges against a single point for  $t \rightarrow \infty$ .
- The system is **unstable** if there is at least one solution that diverges from the stability region permanently for  $t \rightarrow \infty$ .
- The system is **(neutrally) stable** or stable if nearby trajectories remain nearby for  $t \rightarrow \infty$ .
- Obviously, these properties depend on the eigenvalues of the matrix **A**.

# Stable and unstable subspaces I

- We have shown that individual eigenvalues (and thus dynamical ‘modes’) of the system are associated with independent linear subspaces.
- Let  $\mathbf{w}_j = \mathbf{u}_j + i \mathbf{v}_j$  be the generalized complex eigenvector of the (real) matrix  $\mathbf{A}$  belonging to eigenvalue  $\lambda_j = a_j + ib_j$ .
- We define the following linear subspaces:

**stable subspace:**  $E^s = \text{Span}\{\mathbf{u}_j, \mathbf{v}_j \mid a_j < 0\}$

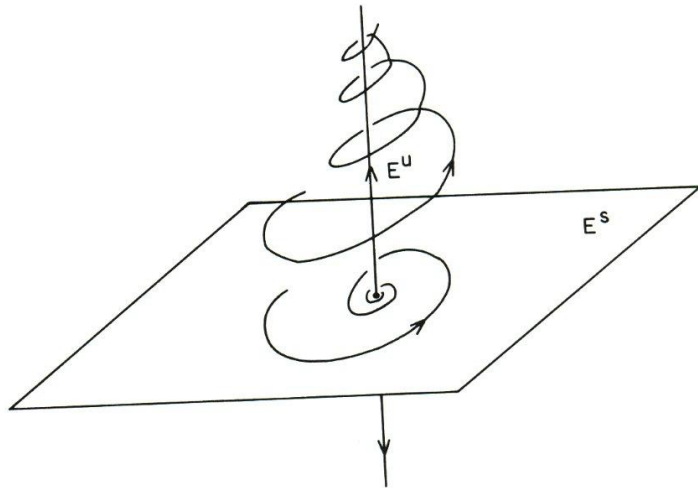
**center subspace:**  $E^c = \text{Span}\{\mathbf{u}_j, \mathbf{v}_j \mid a_j = 0\}$

**unstable subspace:**  $E^u = \text{Span}\{\mathbf{u}_j, \mathbf{v}_j \mid a_j > 0\}$

The stable / unstable / neutrally stable modes of the solution evolve separately in these subspaces.

# Stable and unstable subspaces II

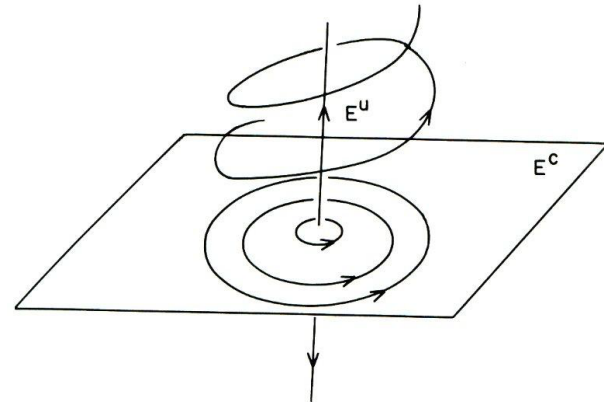
- Examples:



$$A = \begin{bmatrix} -2 & -1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$w_{1,2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \pm i \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{belonging to } \lambda_{1,2} = -2 \pm i$$

$$w_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{belonging to } \lambda_3 = 3$$



$$A = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$w_{1,2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \pm i \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{belonging to } \lambda_{1,2} = \pm i$$

$$w_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{belonging to } \lambda_3 = 2$$

# Invariance of the subspaces

- It can be shown that these subspaces are invariant under the flow  $\varphi(t, \mathbf{x}_0) = e^{t\mathbf{A}} \mathbf{x}_0$ ; this means a solution that starts in a subspace  $E^i$  remains in this space forever.
- In addition, for an  $n \times n$  matrix  $\mathbf{A}$  the subspaces span the whole  $\mathbb{R}^n$ :

$$\mathbb{R}^n = E^s \oplus E^c \oplus E^u$$

# Example: linear recurrent network I

Recurrent weights

Feed-forward weights

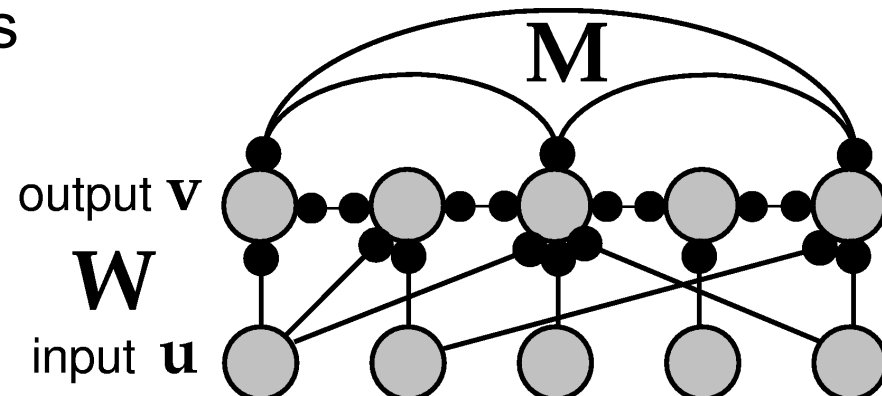
- Mean firing rate approximation results in the DEQ:

$$\tau \dot{\mathbf{v}} = -\mathbf{v} + \mathbf{M}\mathbf{v} + \mathbf{s} \quad \text{with } \mathbf{s} = \mathbf{W}\mathbf{u}.$$

- The system matrix  $\mathbf{A} = \mathbf{M} - \mathbf{I}$  can be diagonalized; we assume that  $\mathbf{M}$  is symmetric, implying that  $\mathbf{M} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$  with  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_N)$  and  $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ .
- Transforming the system to new coordinates  $\mathbf{z}(t) = \mathbf{Q}^T \mathbf{v}(t)$  we obtain the system:  $\tau \dot{\mathbf{z}} = -\mathbf{z} + \mathbf{\Lambda} \mathbf{z} + \mathbf{Q}^T \mathbf{s}$

- When  $\mathbf{q}_n$  signifies the columns of  $\mathbf{Q}$  this implies the decoupled set of equations:

$$\tau \dot{z}_n = -z_n + \lambda_n z_n + \mathbf{q}_n^T \mathbf{s}$$



# Example: linear recurrent network II

- The solution of these DEQs is:

$$z_n(t) = \frac{\mathbf{q}_n^T \mathbf{s}}{1 - \lambda_n} \left( 1 - \exp\left(-\frac{1 - \lambda_n}{\tau} t\right) \right) + z_n(0) \exp\left(-\frac{1 - \lambda_n}{\tau} t\right)$$

- The stationary solution in original coordinates is thus:

$$\mathbf{v}(\infty) = \mathbf{Q}\mathbf{z}(\infty) = \sum_{n=1}^N \frac{\mathbf{q}_n^T \mathbf{s}}{1 - \lambda_n} \mathbf{q}_n \quad (\text{assuming } \lambda_n < 1)$$

- Dynamics defined by linear combination of independent ‘modes’, driven by projections of input signal onto the eigenvectors of  $\mathbf{M}$ .
- If one eigenvalue is close to one, and not the others, the corresponding eigenmode is selectively amplified:  
 $\mathbf{v}(\infty) \approx \frac{\mathbf{q}_1^T \mathbf{s}}{1 - \lambda_1} \mathbf{q}_1$ . Similarly, multiple eigencomponents can be amplified.

# Example: linear recurrent network III

- If  $\lambda_1 = 1$  and  $\lambda_n \ll 1$  for  $n > 1$ , the network can act as an **integrator** since then holds:  $\tau \dot{z}_1 = \mathbf{q}_1^T \mathbf{s}$

Solution (even for time-dependent input):

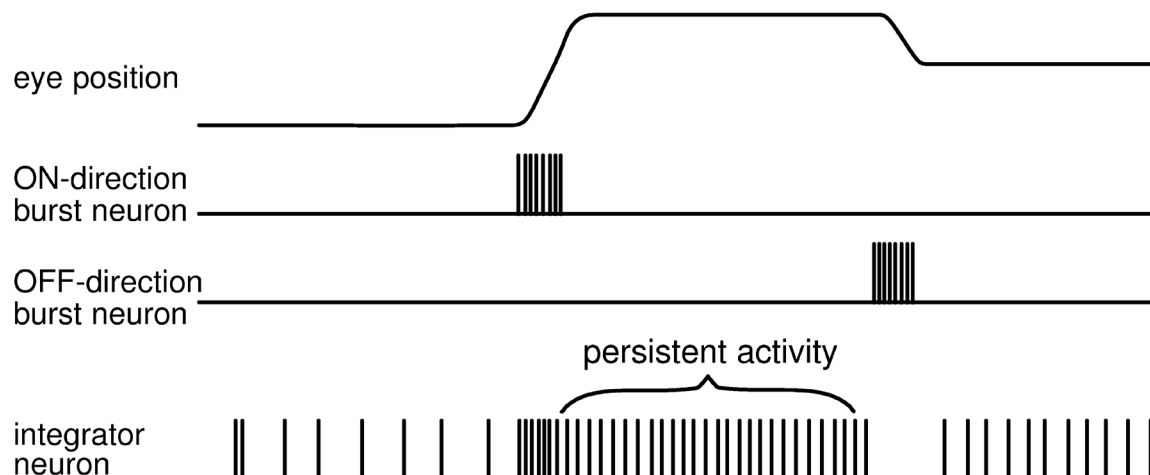
$$z_1(t) = z_1(0) + \frac{1}{\tau} \int_0^t \mathbf{q}_1^T \mathbf{s}(t') dt'$$

With  $\lambda_1 \gg \lambda_n$  follows:

$$\mathbf{v}(t) \approx \frac{\mathbf{q}_1}{\tau} \int_0^t \mathbf{q}_1^T \mathbf{s}(t') dt'$$

(for sufficiently large times, so that the modes for the  $\lambda_n \ll 1$  modes have decayed to zero)

- Neurons with such integrator characteristics have been found in the brain stem in networks that control eye position.



# Things to remember

- Continuous vs. discrete DS  $\rightarrow 2)$
- Autonomous DS  $\rightarrow 2)$
- Linear DS  $\rightarrow 2)$
- Exponential of a matrix  $\rightarrow 2)$
- Phase portrait  $\rightarrow 2)$
- General form of the solution for linear autonomous DS  $\rightarrow 2)$
- Stability-related subspaces  $\rightarrow 2)$
- Neural network applications  $\rightarrow 1,3)$



# Literature (for this lecture)

- 1) Dayan, P. & Abbott, L.F. (2001 / 2005) *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, Cambridge MA, USA. Chapter 7.
- 2) Perko, L. (1998) *Differential Equations and Dynamical Systems*. Springer-Verlag, Berlin. Chapter 1.
- 3) Wilson, H.R. (1999) *Spikes, Decisions, and Actions*. Oxford University Press, UK. Chapters 2-4. (Can be downloaded from Hugh Wilson's homepage !)