Recurrent Neural Networks and Dynamical System

Neuro 120, Section 8

Brief Conceptual Review and HW4

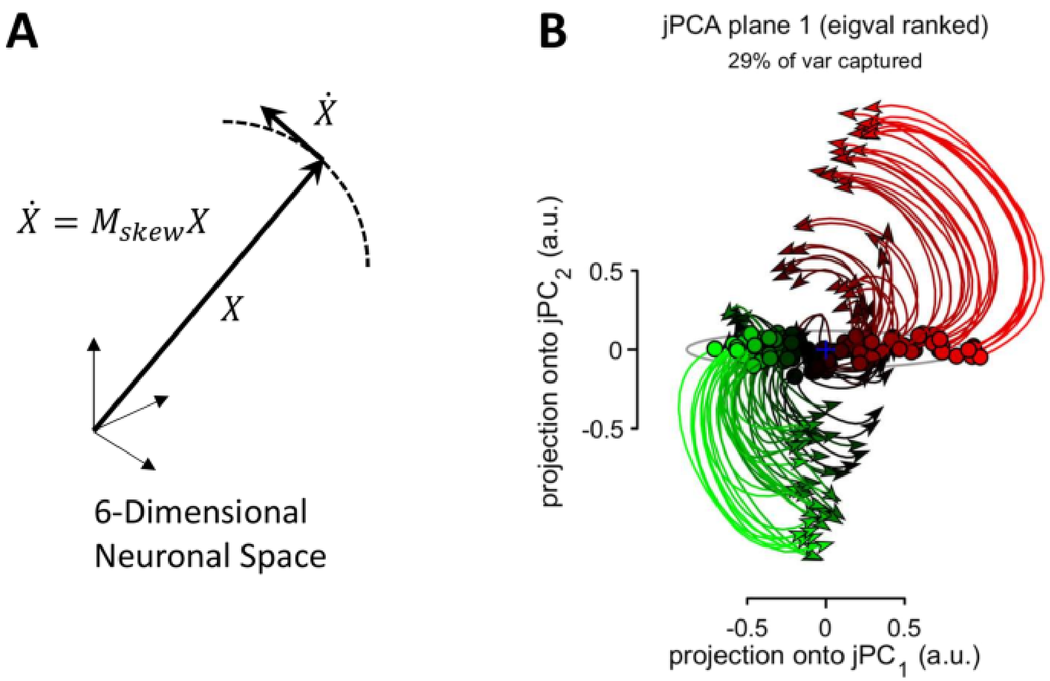
Binxu Wang

Why we care about dynamics in Neuroscience?

- Brain operates in time.

- Dynamics could implement many (cognitive) computation!

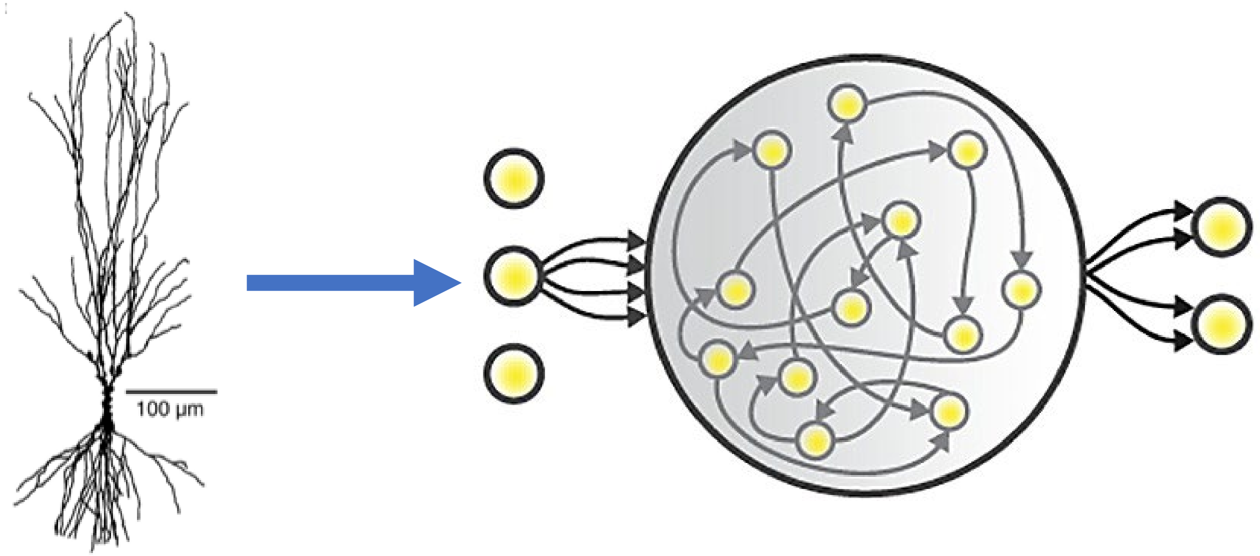
  - **Memory**
  - **Information integration**
  - Decision making
  - Pattern completion… (Hopfield)



Analysis of neuronal ensemble activity reveals the pitfalls and shortcomings of rotation dynamics

Choice to make when building neural networks

- Models of Neurons (Units)
    - Type of output: Rate-based vs Spike-based
    - LIF, FN, HH model …
    - Compartmentation
    - Rate equation



- Models of Synaptic Connection
    - Scaler weight (Excitatory vs Inhibitory) / Temporal kernel
    - Plasticity / learning rule

- Models of Network Structure
    - All-to-all Fully connected
    - Random matrix
    - Sparse
    - Spatially regular structure (local, CNN)
    - Learned structure

Main Neural Network Model

- Nonlinear version

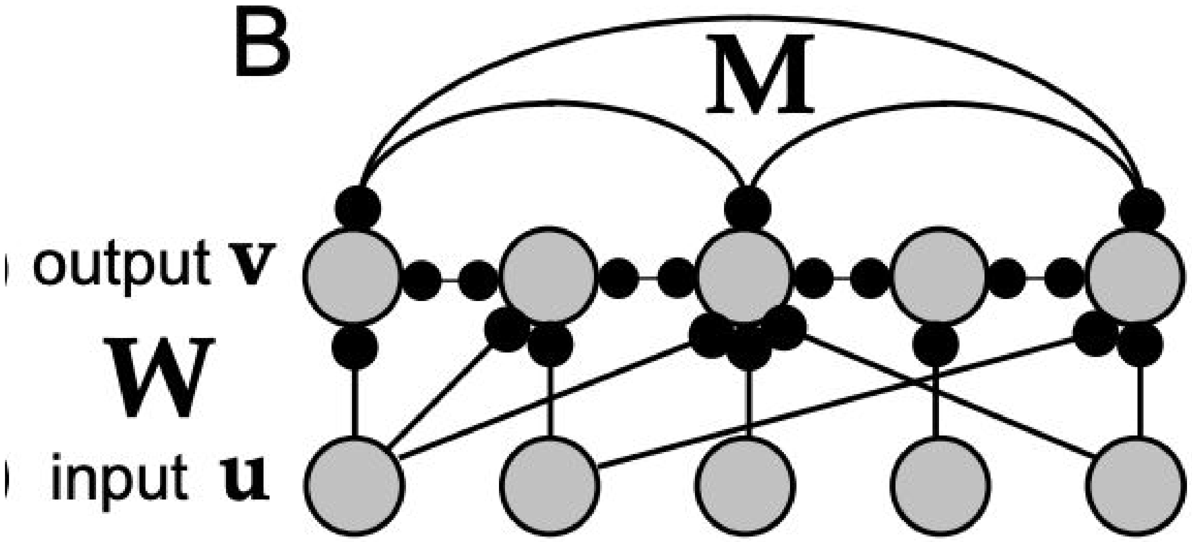
$$\tau \frac{dv}{dt} = -v + F(Mv + Wu)$$

  - What are the meaning of these components?

- Linear version

$$\tau \frac{dv}{dt} = -v + Mv + Wu$$

  - $h = Wu$



Linear Dynamic Systems Recep (II)

Why we need good intuitions for the dynamical system analysis?

- As computational neuroscientist, our job is not just to analyze existing models, but to design new ones!

$$\text{Connectivity} \Leftrightarrow \text{Dynamics}$$

  - We need to “guess” the underlying network structure from the observed dynamics, based on biology.
  - Or build model to accomplish a desired computation.

- Linear system analysis is the first step to build such intuition.

Dynamic System Review

$$\dot{X} = f(X)$$

X is state variables

$f(X)$ defines the dynamics or motion.

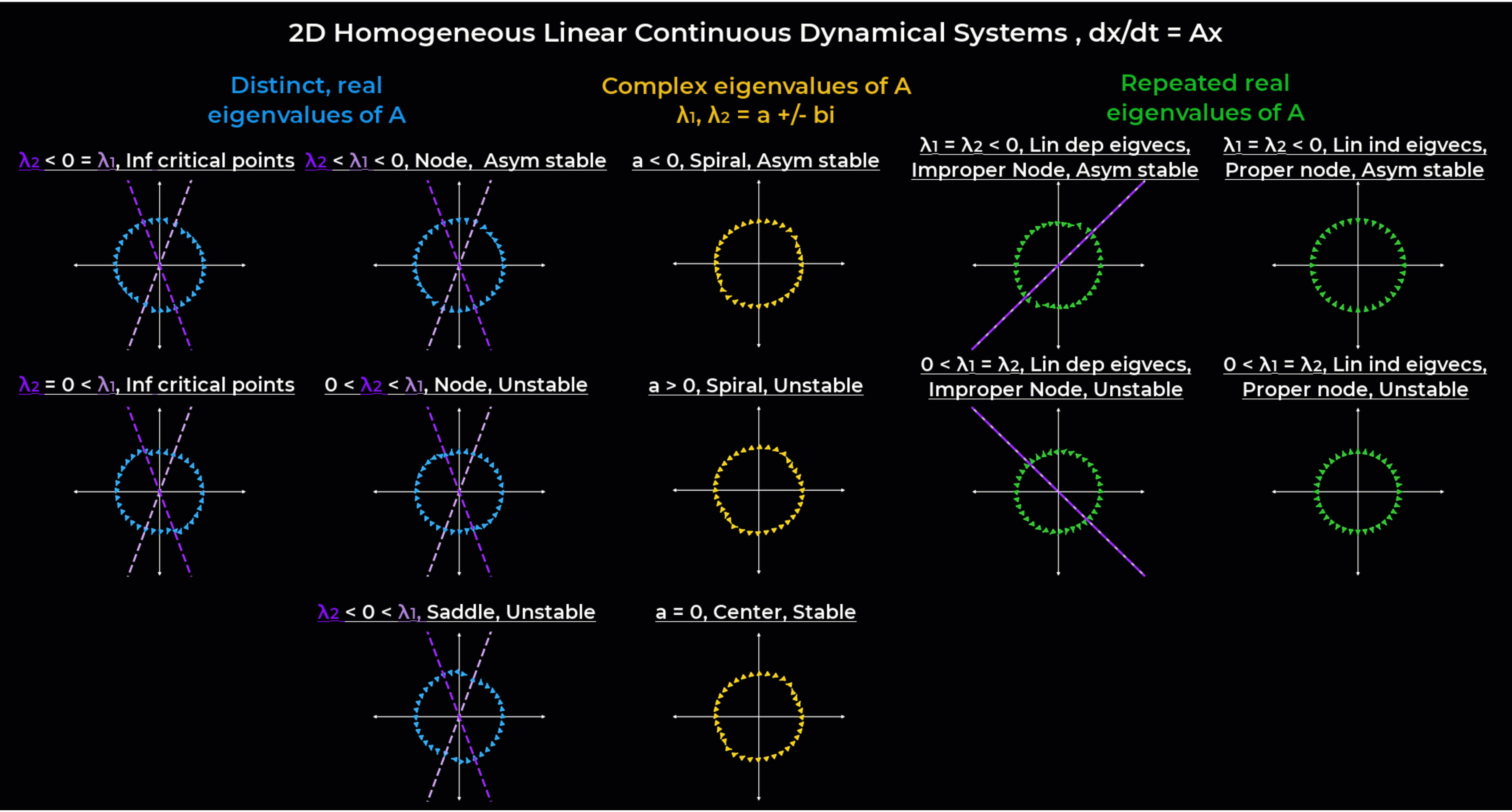
Analysis View

$$f(X) = \dot{X} \approx \frac{\Delta X}{\Delta t}$$
$$\Delta X \approx f(X)\Delta \mathrm{t}$$
$$X(t + \Delta t) \approx X(t) + f(X)\Delta \mathrm{t}$$

- ~ Euler method for integrating / solving ODE. (HW3, P2)
- A few ODE has analytical solutions $X(t)$
  - $X(t) = e^{kt}$, then $\dot{X} = ke^{kt} = kX$
  - $X(t) = e^{tA}X(0)$ then $\dot{X} = Ae^{tA}X(0) = AX(t)$

Geometric View

- Differential equations define a vector field.
  - $f(X)$ is the vector at X
- Differential equations / Vector field can be integrated as a flow $X(t)$.
- Dynamics could be read out from the geometry of flow field.
  - Fixed points (attractor, repeller, saddle etc.)
  - Limit cycle


2D Homogeneous Linear Continuous Dynamical Systems , dx/dt = Ax
Distinct, real eigenvalues of A
Complex eigenvalues of A
$\lambda_1, \lambda_2 = a +/- bi$
Repeated real eigenvalues of A
$\lambda_2 < 0 = \lambda_1$, Inf critical points
$\lambda_2 < \lambda_1 < 0$, Node, Asym stable
a < 0, Spiral, Asym stable
$\lambda_1 = \lambda_2 < 0$, Lin dep eigvecs, Improper Node, Asym stable
$\lambda_1 = \lambda_2 < 0$, Lin ind eigvecs, Proper node, Asym stable
$\lambda_2 = 0 < \lambda_1$, Inf critical points
$0 < \lambda_2 < \lambda_1$, Node, Unstable
a > 0, Spiral, Unstable
$0 < \lambda_1 = \lambda_2$, Lin dep eigvecs, Improper Node, Unstable
$0 < \lambda_1 = \lambda_2$, Lin ind eigvecs, Proper node, Unstable
$\lambda_2 < 0 < \lambda_1$, Saddle, Unstable
a = 0, Center, Stable

Eigenvalues of Real Matrix

- Eigenvalue problem: For $n \times n$ matrix M

$$Mv = \lambda v$$

  - Usually allowed λ, v to be complex

- When M has n *linear independent* eigenvectors
  $\Leftrightarrow$ M is diagonalizable.

$$V = [v_1, \dots v_n]$$
$$\Lambda = diag([\lambda_1, \dots \lambda_n])$$

  - Then we can perform eigen decomposition of M

$$MV = V\Lambda$$
$$M = V\Lambda V^{-1}$$

Special Case: Symmetric matrix

- Real symmetric matrix M

$$M = M^T$$
$$M_{ij} = M_{ji}$$

  - What that means for a network connectivity matrix? Realizable?
    - Exact reciprocal connection.

- Eigenvalues
  - All eigenvalues λ_i of M are real.

- Eigenvectors
  - All eigenvectors of M form orthonormal basis.
  - i.e. V is orthogonal matrix. $VV^T = V^T V = I$

$$M = V\Lambda V^{-1} = V\Lambda V^T$$

What kind of dynamics symmetric matrices cannot support ?

- No Spiral or rotational dynamics!
- Hopfield network? Feels similar

Special Case: Eigens of this Matrix

- *What are the Eigenvectors of this matrix?*

$$\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

```
julia> Dmat = Float64[3 1;0.000 3]
2x2 Matrix{Float64}:
 3.0  1.0
 0.0  3.0

julia> eva = eigvals(Dmat)
2-element Vector{Float64}:
 3.0
 3.0

julia> evc = eigvecs(Dmat)
2x2 Matrix{Float64}:
 1.0  -1.0
 0.0   6.66134e-16
```

```
julia> evc*diagm(eva)*inv(evc)
2x2 Matrix{Float64}:
 3.0  0.0
 0.0  3.0
```

What??? Is algebra broken???

```
julia> inv(evc)
2x2 Matrix{Float64}:
 1.0  1.5012e15
 0.0  1.5012e15
```

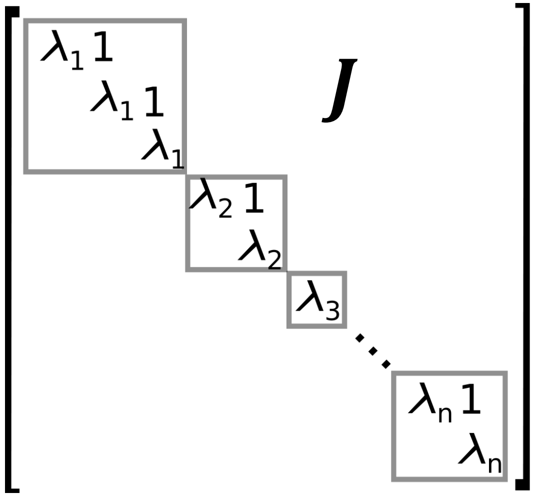
Not every real matrix is diagonalizable

- Some matrices don't have n linearly independent eigenvectors.
  - So, you cannot do decomposition $M = V\Lambda V^{-1}$ because you cannot do inversion (reliably…).
  - ~Jordon canonical form
    - $A = PJP^{-1}$, J looks like

- BUT, (on computer) you can always find eigenstuffs for a real $n \times n$ matrix M.

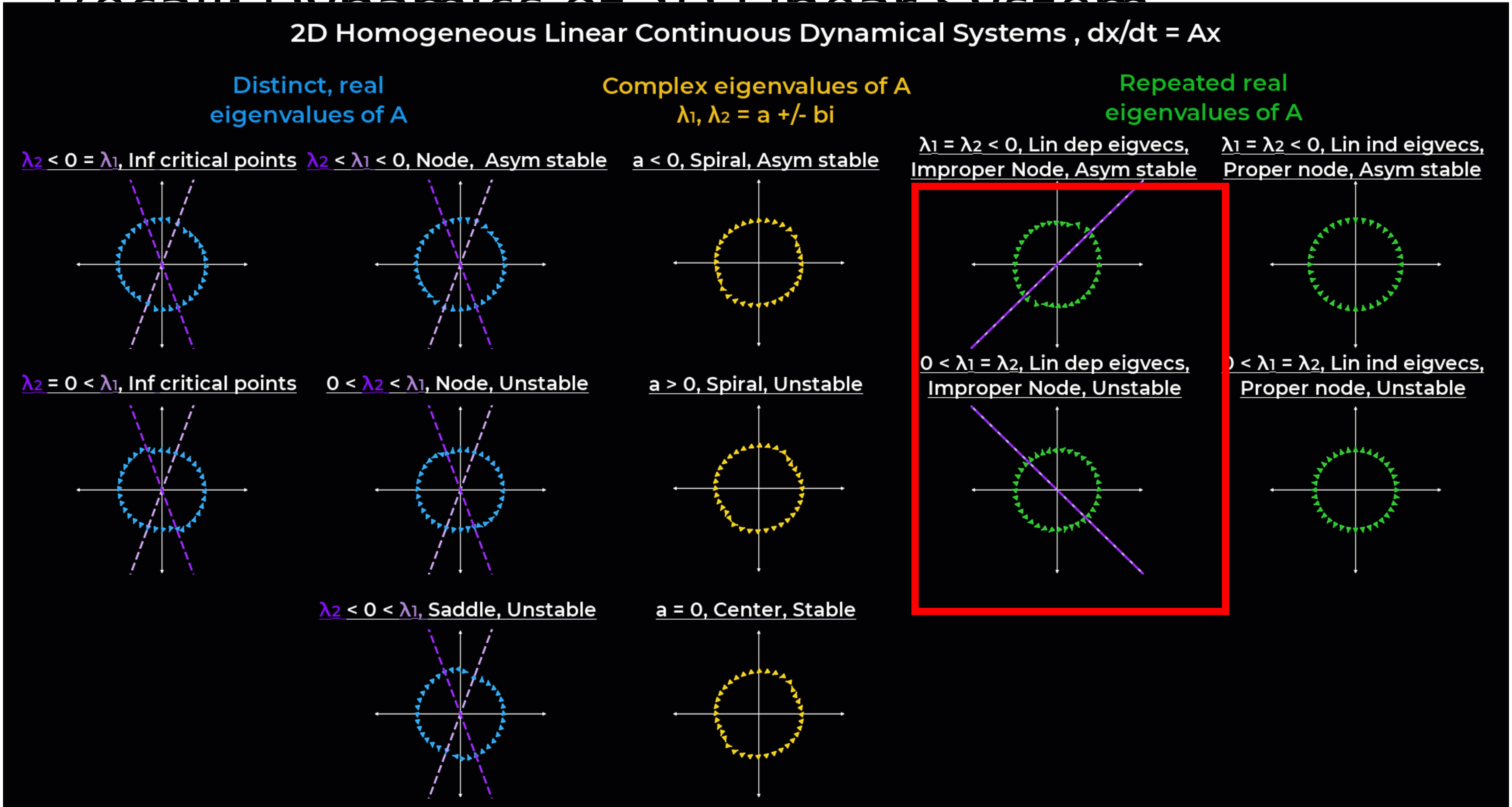
```
julia> eva = eigvals(Dmat)
2-element Vector{Float64}:
 3.0
 3.0

julia> evc = eigvecs(Dmat)
2×2 Matrix{Float64}:
 1.0  -1.0
 0.0   6.66134e-16
```

```
julia> inv(evc)
2×2 Matrix{Float64}:
 1.0  1.5012e15
 0.0  1.5012e15
```



[You cannot compute the Jordan canonical form numerically?]

Recall: Dynamics of 2D Linear System



General Solution to Linear System

- **Assume A is diagonalizable**
  - $A = P\Lambda P^{-1}$
- Solution of linear system

$$\frac{dv}{dt} = Av$$

  - Is the matrix exponential

$$v(t) = e^{tA}v(0) = Pe^{t\Lambda}P^{-1}v(0)$$

$$e^{t\Lambda} = \begin{bmatrix} e^{t\lambda_1} & 0 & 0 \\ 0 & e^{t\lambda_2} & 0 \\ 0 & 0 & e^{t\lambda_n} \end{bmatrix}$$

- Basically,
  - transform initial state onto eigenbasis P;
  - each mode evolve exponentially in independence;
  - change basis back

If we change to the new basis, then the new dynamic system will be extremely simple

$$\tilde{v} = P^{-1}v$$

- The dynamics will be orthogonal

$$\frac{d\tilde{v}}{dt} = \Lambda\tilde{v}$$

$$\tilde{v}(t) = e^{t\Lambda}\tilde{v}(0)$$

Attention: Orthonormal vs Non-orthonormal Basis

Non-orthonormal Basis

- Property
  - $V^T V \neq I \neq V V^T$
- Basis change requires matrix inversion:
  - $c = V^{-1} v$
  - not just inner product $c_i \neq v_i^T v(0)$

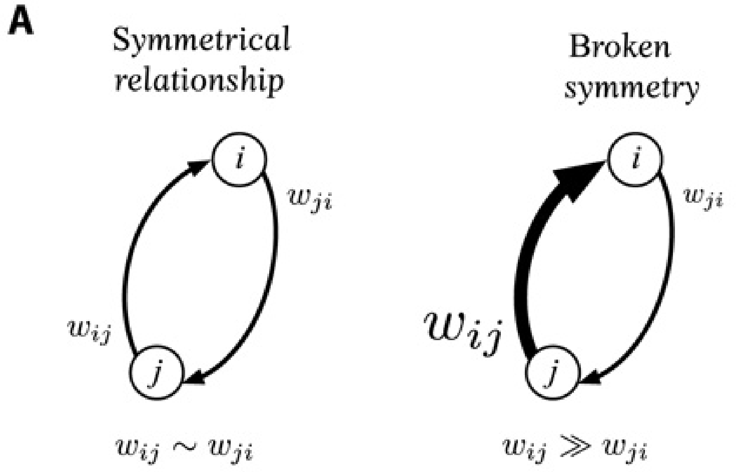
Orthonormal Basis

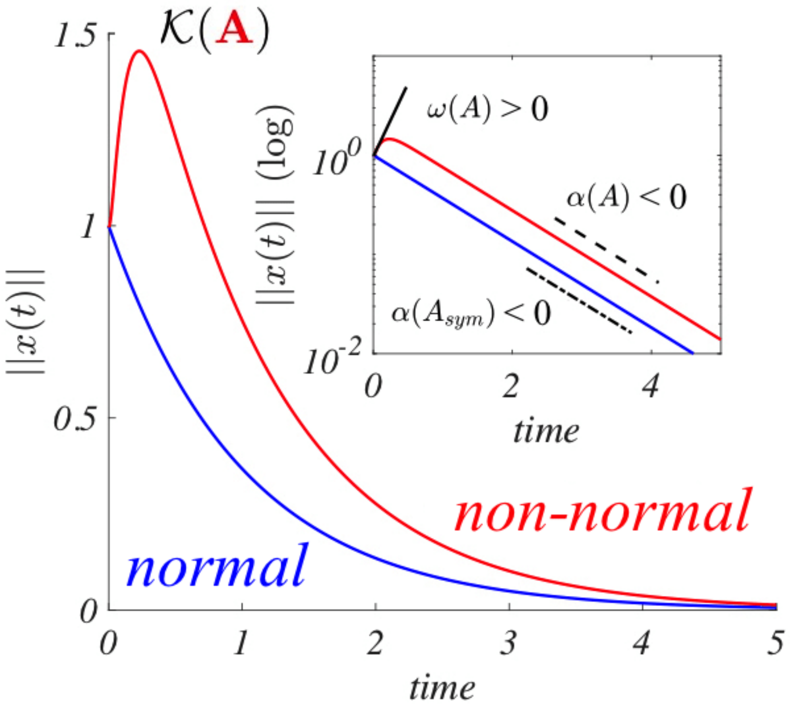
- Property
  - $V V^T = I = V^T V$
  - $V^T = V^{-1}$
- Basis change by inner product:
  - $c = V^T v$
  - $c_i = v_i^T v$
- Special cases:
  - symmetric or skew symmetric matrix …
  - Basis we got in PCA, SVD etc.

Non-normal Dynamics is Interesting

- Eigenvectors do not form orthonormal basis!

- Highly non-symmetric interaction creates non-normal dynamics ⇒ Large transients, which explain some cortical dynamical phenomenon.





[Non-normal Recurrent Neural Network (nnRNN): learning long time dependencies while improving expressivity with transient dynamics]()

[**Structure and dynamical behavior of non-normal networks**]()

[**Non-normal amplification in random balanced neuronal networks**]()

[Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns]()

Linear Dynamic System: Eigen-modes

$$\frac{dv}{dt} = Av$$

- What if v is along an eigenvector of A?

$$A\boldsymbol{v_i} = \lambda_i \boldsymbol{v_i}$$
$$v(0) = c(0)\boldsymbol{v_i}$$

  - What will be the dynamics?
  - Change to the state v always go along v

$$\frac{dv}{dt} = \lambda_i v$$
$$\frac{dc}{dt}\boldsymbol{v_i} = \lambda_i c \boldsymbol{v_i}$$
$$v(t) = v(0)e^{\lambda_i t}$$

Dynamics depend on λ

- $\lambda_i < 0$
  - Exponential decay to zero
- $\lambda_i = 0$
  - Stay at $v(0)$, stay put.
- $\lambda_i > 0$
  - Exponential explode

Linear Dynamic System: Eigen-modes

$$\frac{dv}{dt} = Av$$

- What if v starts as a combination of (linear independent) eigen vectors?

$$A\boldsymbol{v_i} = \lambda_i \boldsymbol{v_i}$$
$$v(0) = c_1(0)\boldsymbol{v_1} + c_2(0)\boldsymbol{v_2}; \lambda_1, \lambda_2 \in \mathbb{R}$$

- What will be the dynamics?
  - Dynamics along each eigenvector operate on its own.

$$v(t) = c_1(0)\boldsymbol{v_1}e^{\lambda_1 t} + c_2(0)\boldsymbol{v_2}e^{\lambda_2 t}$$

$$\frac{dc_1}{dt}\boldsymbol{v_1} + \frac{dc_2}{dt}\boldsymbol{v_2} = A(c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2})$$
$$\frac{dc_1}{dt}\boldsymbol{v_1} + \frac{dc_2}{dt}\boldsymbol{v_2} = c_1\lambda_1\boldsymbol{v_1} + c_2\lambda_2\boldsymbol{v_2}$$

Bases on definition of linear independence, there is only one way of decomposing a vector. Thus, the coefficient must match.

$$\frac{dc_1}{dt} = c_1\lambda_1$$
$$\frac{dc_2}{dt} = c_2\lambda_2$$

This derivation could be generalized to the full n-dim eigen basis.

Linear Dynamic System: Complex Eigen-modes

$$\frac{dv}{dt} = Av$$

- What if v starts as a combination of complex eigenvectors? (with complex eigenvalues)?

$$A\boldsymbol{v_i} = \lambda_i \boldsymbol{v_i}$$
$$A\overline{\boldsymbol{v}_i} = \overline{\lambda_i}\overline{\boldsymbol{v}_i}$$

$$\boldsymbol{v_i} = \boldsymbol{u} + \boldsymbol{w}\mathrm{i}, \qquad \boldsymbol{v_i} \in \mathbb{C}^n, \boldsymbol{u}, \boldsymbol{w} \in \mathbb{R}^n$$
$$\lambda_i = a + b\mathrm{i}, \qquad \lambda_i \in \mathbb{C}, a, b \in \mathbb{R}$$

- Initial state start in this complex eigen space

$$v(0) = c\boldsymbol{u} + d\boldsymbol{w} = \frac{c - d\mathrm{i}}{2}\boldsymbol{v_i} + \frac{c + d\mathrm{i}}{2}\overline{\boldsymbol{v}_i}$$

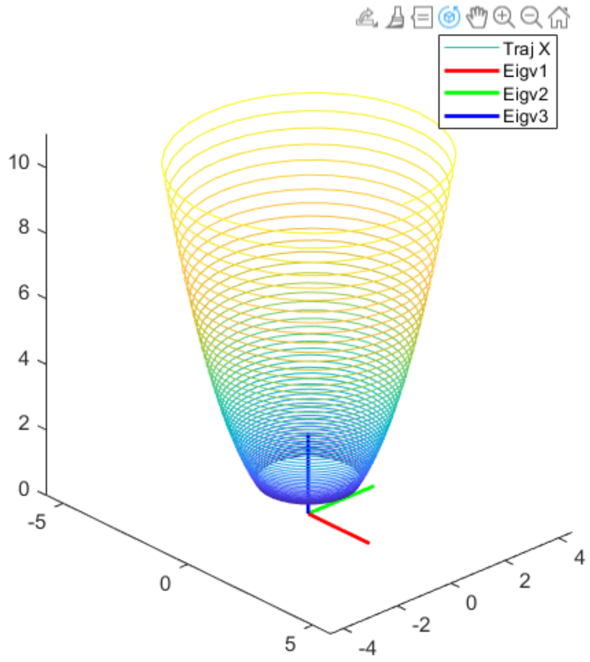
- What will be the dynamics?
  - Rotational dynamics in the space spanned by $\mathrm{Span}(\boldsymbol{u}, \boldsymbol{w})$

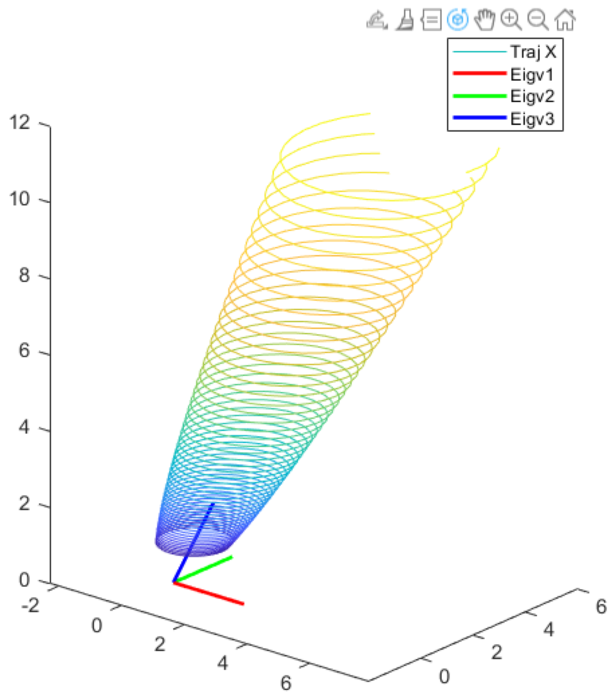
$$v(t) = \frac{c - d\mathrm{i}}{2}\boldsymbol{v_i}e^{(a+b\mathrm{i})t} + \frac{c + d\mathrm{i}}{2}\overline{\boldsymbol{v}_i}e^{(a-b\mathrm{i})t}$$
$$= e^{at}\left(c\,\frac{\boldsymbol{v_i}e^{\mathrm{i}bt} + \overline{\boldsymbol{v}_i}e^{-\mathrm{i}bt}}{2} - di\,\frac{\boldsymbol{v_i}e^{\mathrm{i}bt} - \overline{\boldsymbol{v}_i}e^{-\mathrm{i}bt}}{2}\right)$$
$$= e^{at}[(c\cos bt + d\sin bt)\boldsymbol{u} + (d\cos bt - c\sin bt)\boldsymbol{w}]$$

- Interpretation
  - [ ] is the periodical solution
  - b control the speed or period of rotation
  - a control the amplitude of rotation.

Fun time: How to design a tornado?

- 3D system
- Rotational dynamics
- 1-dim explosion






```
%% Tornado system: Set the kernel (Eigenvalue matrix)
a = +0.035; b = 10.5;
l3 = +0.08;
D = [ a  b   0;
     -b  a   0;
      0  0  l3];
%
basis = [1  0  0;
         0  1  0;
         0  0  1];
basis = basis ./ sqrt(sum(basis.^2,1));
%% Set dynamic matrix and run!
MTor = basis * D * inv(basis);

X0 = [1,1,1];
odesol = ode45(@(t,X)MTor*X,[0 30],X0,odeset('RelTol',1E-6));
Figure;hold on
patch([odesol.y(1,:),nan],[odesol.y(2,:),nan],[odesol.y(3,:),nan],[odesol.x,nan],'FaceColor','none','EdgeColor','interp');hold onaxis equal
Scale=set_scale();
plot_basis(basis,Scale)
legend(["Traj X","Eigv1","Eigv2","Eigv3"])
```

Neuro120_tutorials/DSspiral3D_demo.m at master · Animadversio/Neuro120_tutorials

Linear Dynamic System: Fixed point(s)

$$\frac{dv}{dt} = Av$$

- How many steady state could a linear dynamic system possibly have?
  - Two and only two?
  - One?
  - Infinite?
- If more than one, are they connected?
  - Yes, by linearity.

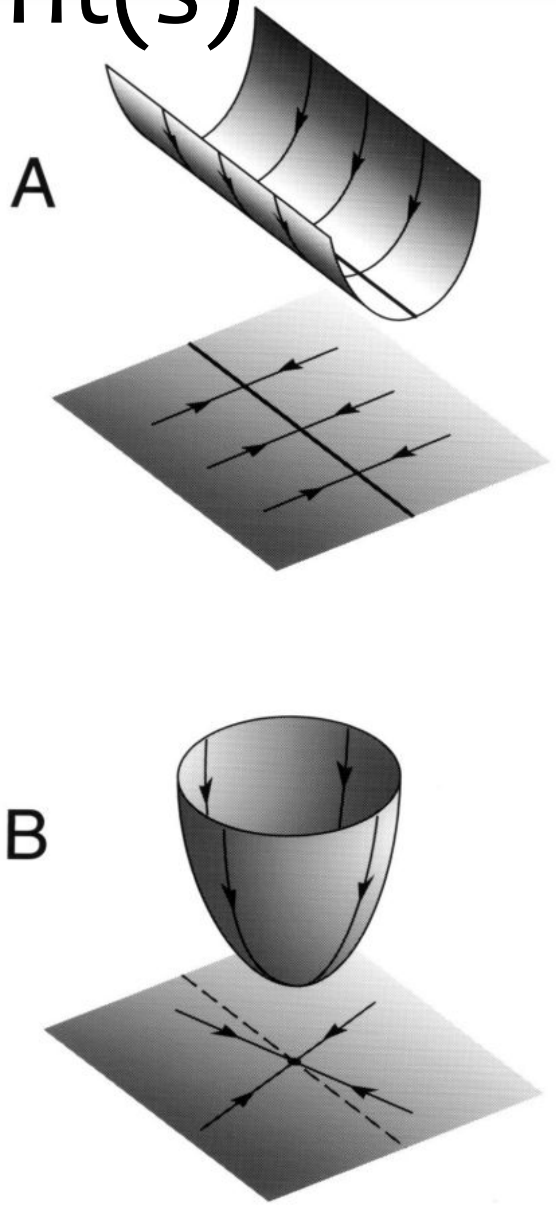
Set of fix points are the solution to linear equation.

$$\{v \in \mathbb{R}^n | Av = 0\}$$

- $v = 0$ always
- $v \in Span\{v_i\}$, eigenvector space with 0 eigenvalue. i.e. $\ker(A)$

Linear Dynamic System: Fixed point(s)

- What's the dimensionality of the space of fixed points?
  - Dimensionality of the kernel space of A
  - Number of eigenvector with $\lambda = 0$
- Why this dimensionality matters?
  - ~ Continuous Attractor (Line/Plane/Ring)
  - Functionally, ~ the dimensionality of variable you want to encode.
    - Head direction ~ 1D (or 2D?)
    - Location ~ 2D



Linear Dynamic System: Fixed point stability

- Under what criterion a fixed point is stable?
  - For converging: All real part of eigenvalue $\mathrm{Re}\lambda < 0$
  - For not diverging : All real part of eigenvalue $\mathrm{Re}\lambda \leq 0$

Linear System: with Inputs

$$\frac{dv}{dt} = Av + g$$

- Fixed point(s)
  - Set of fix points: $\{v|Av = -g\}$
  - Dimensionality: dim of kernel space of A.

- Dynamics around fixed point
  - Let $Av^* + g = 0$, then the equation becomes
  $$\frac{d(v - v^*)}{dt} = A(v - v^*)$$
  - Same linear dynamics around a different origin.
  $$\frac{d\tilde{v}}{dt} = A\tilde{v}$$

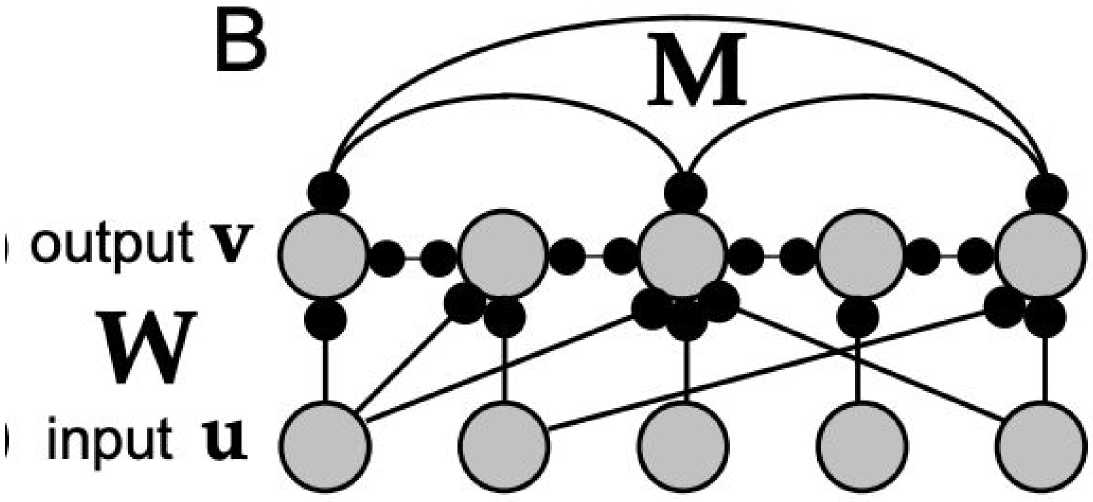
How this general theory applies to Linear RNNs?

$$\tau \frac{dv}{dt} = -v + Mv + h$$

- We can translate previous results by letting

$$A = \frac{1}{\tau}(M - I)$$
$$g = \frac{1}{\tau}h$$

  - How this affects eigen vectors?
    - No change
  - How this affects eigen values?
    - Translate the eigenvalues
    - $\lambda_i(A) = \lambda_i(M) - 1$
    - $M = V\Lambda V^{-1}, A = V(\Lambda - I)V^{-1}$



How to design a neural integrator ?

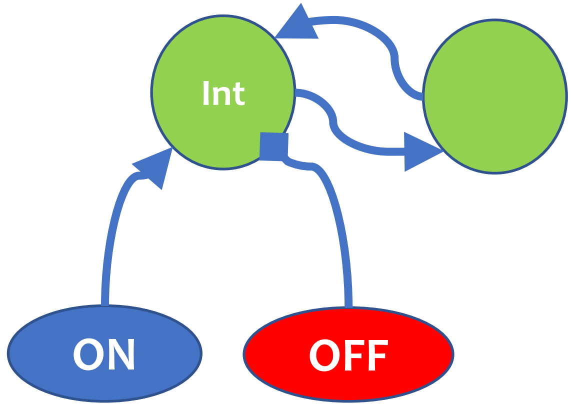
- Input $u(t)$

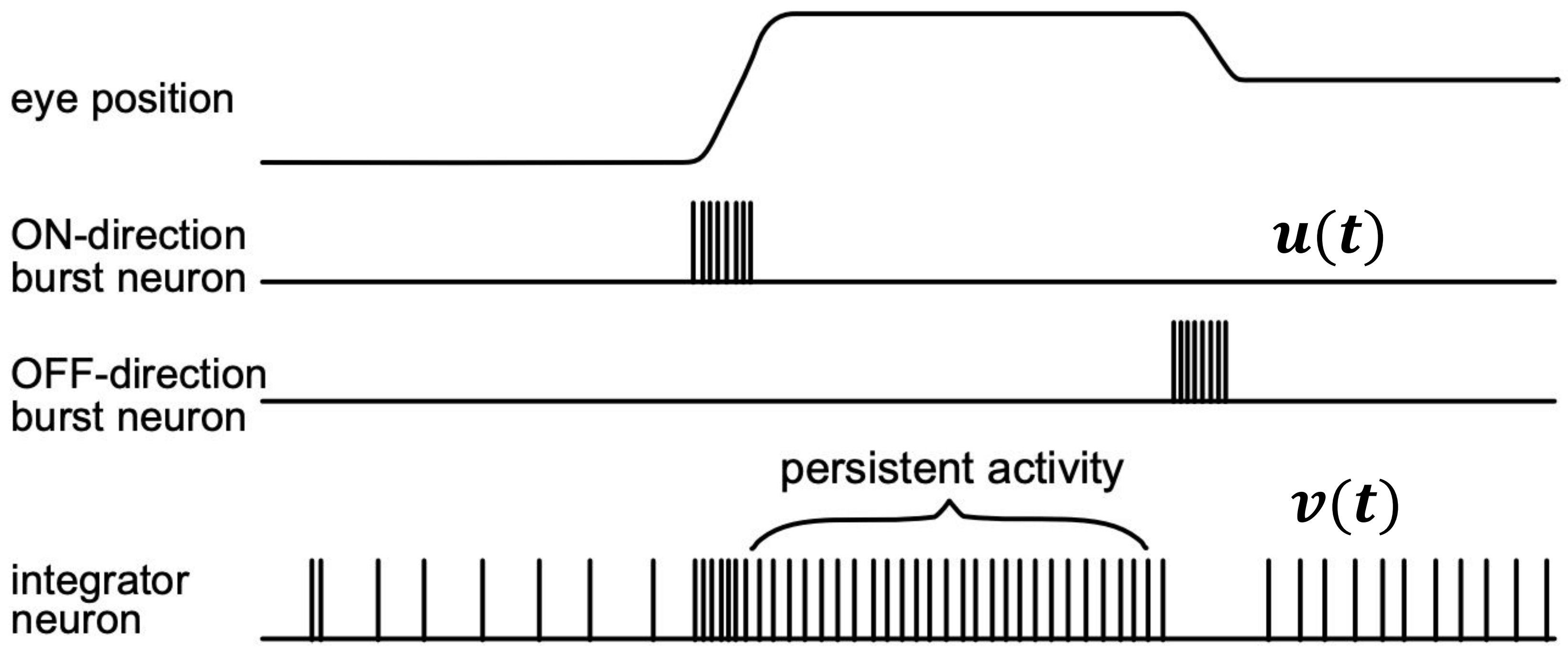
- Output with dynamic
  - $\frac{dv}{dt} = u(t)$

- Recurrent part has 0 eigenvalue, $(-v + Mv) = 0$

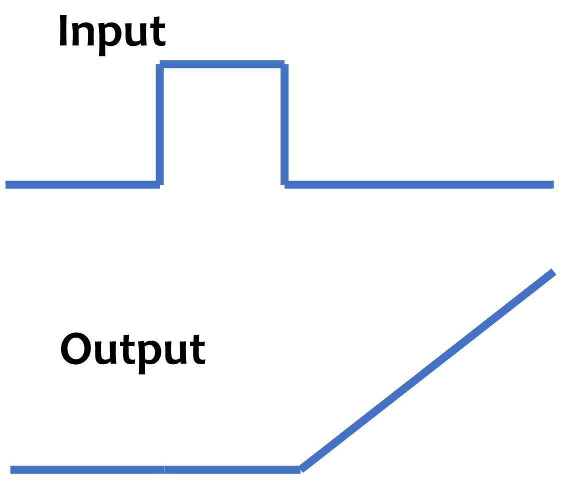
Example: Neural Integrator, eye position memory.

- Recurrent neurons maintain a 1D line attractor
- ON and OFF neuron push or pull it along the line.

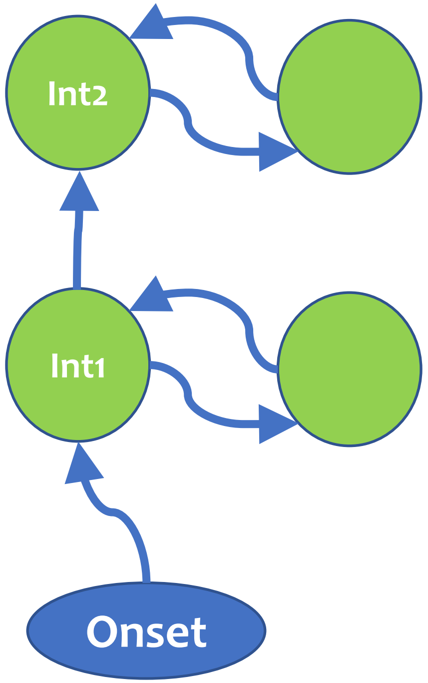




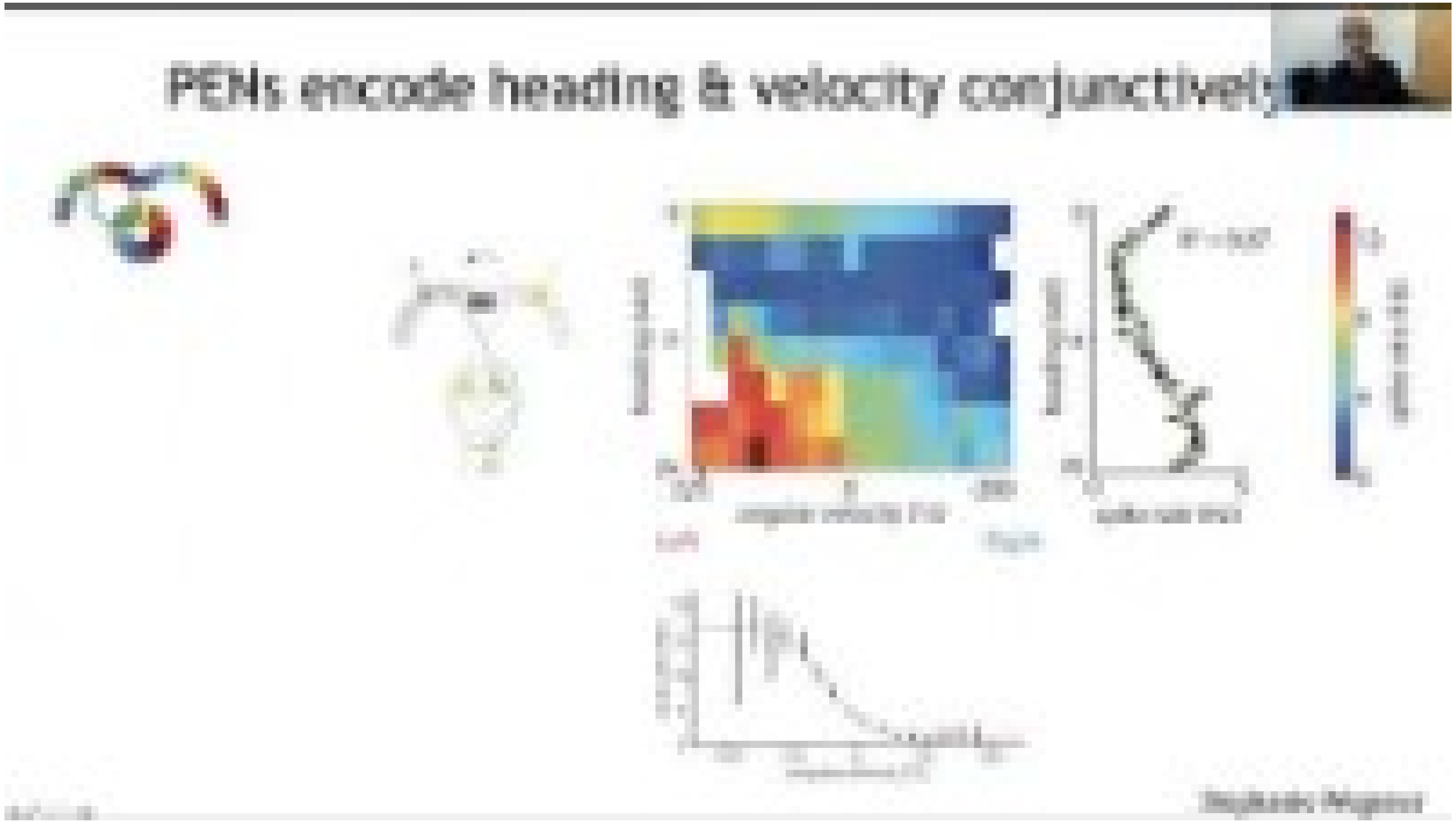
Question: How to build a circuit that have linearly ramping up activity triggered by an impulse?



- One possible solution is to use two integrator modules
  - 1st one integrate the pulse into constant firing
  - 2nd integrator integrate the constant firing into linear ramping activity.



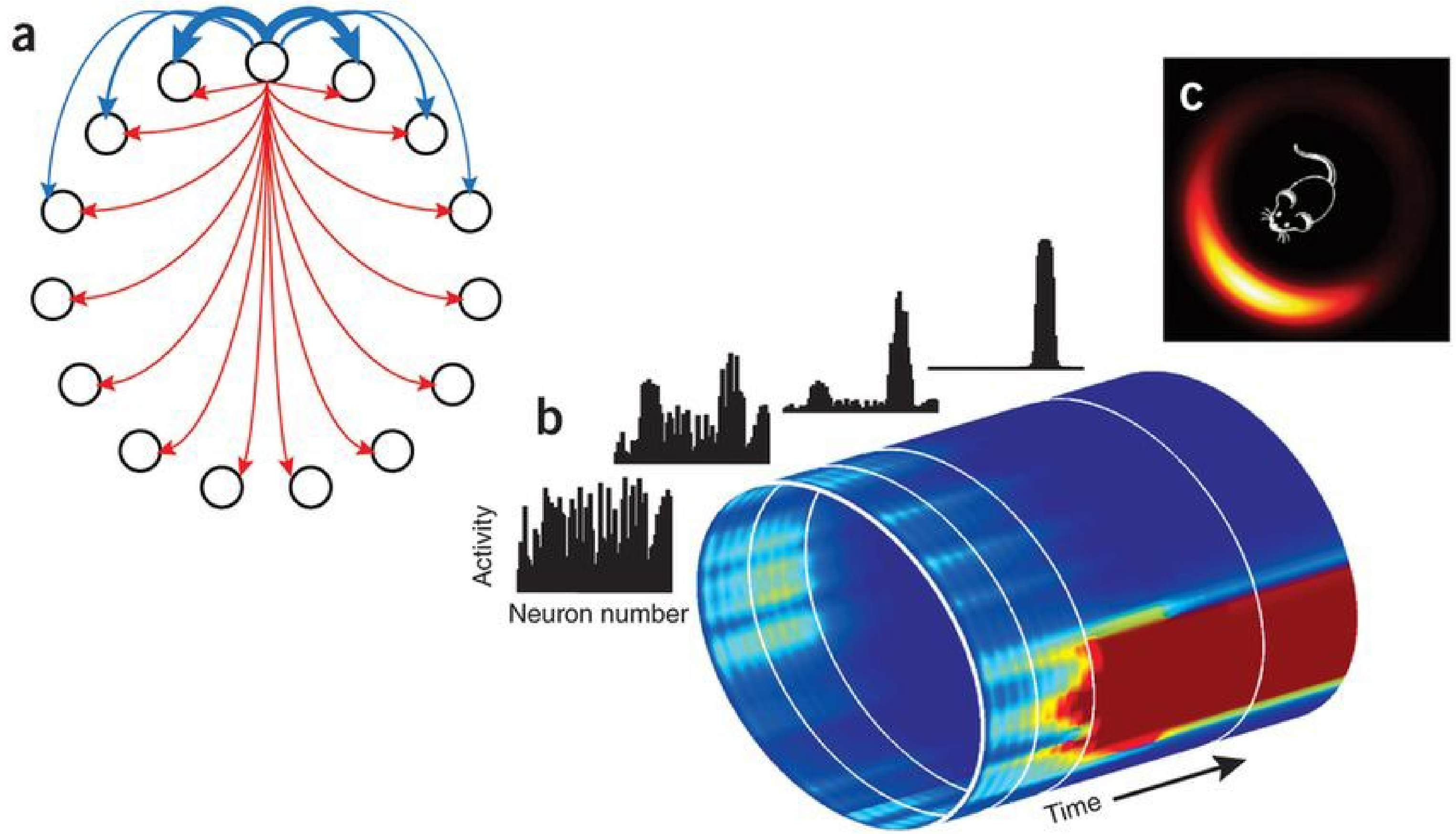
Continuous Ring Attractor for Real in Flys



Can you see the connection to the continuous attractor network?

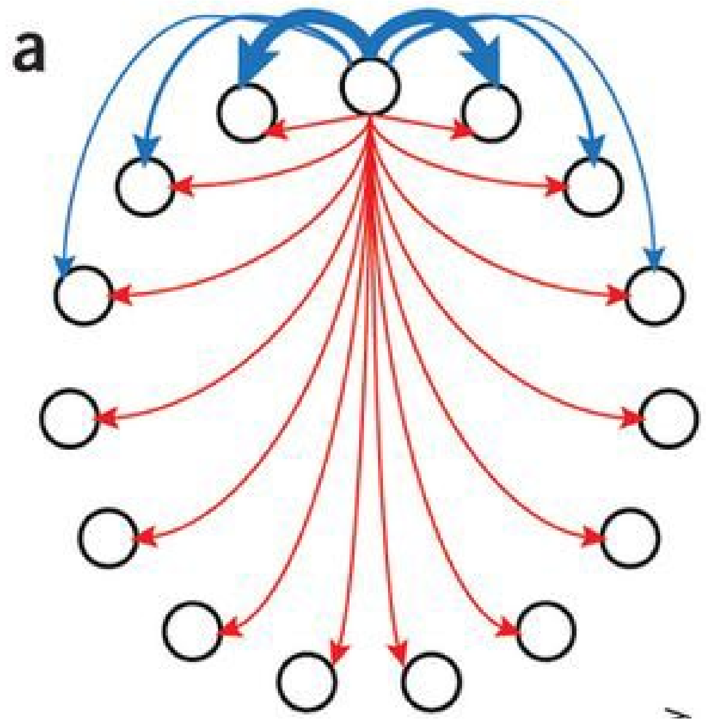
https://youtu.be/2fORQ7icGVU?t=325

Continuous Ring Attractor in Theory



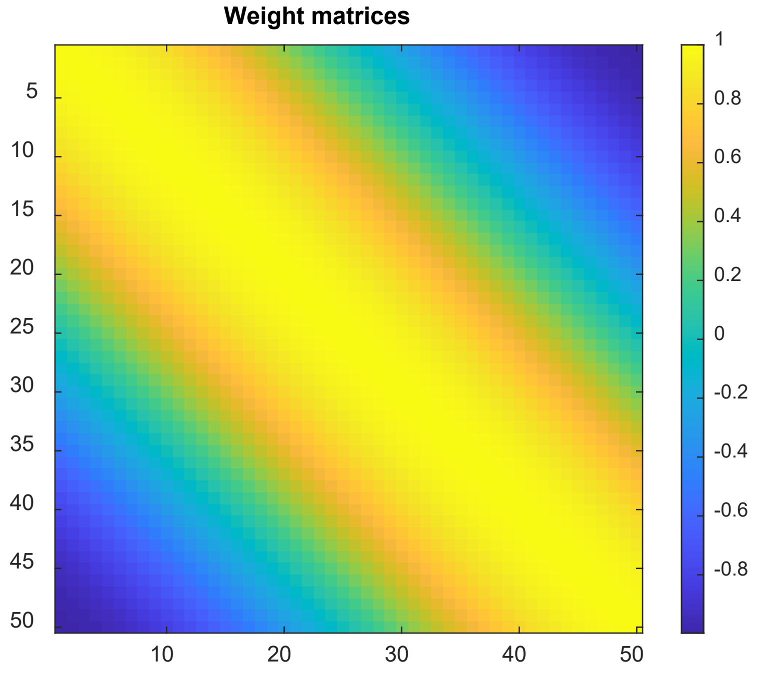
Spectral Analysis of Recurrent Matrix

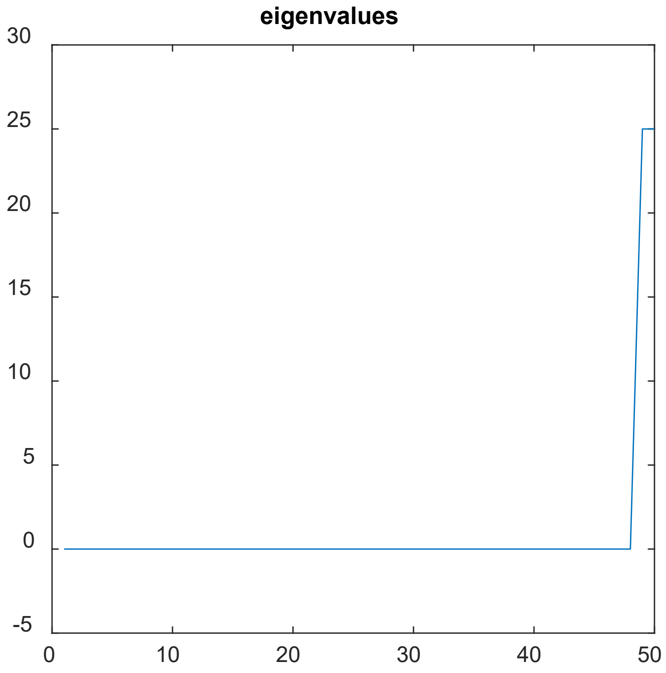
- Recurrent connection matrix
  - $\theta_i = (i-1)\,\Delta\theta$
  - $M[i,j] \propto \cos(\theta_i - \theta_j)$
- All but 2 eigenvalues are zero.
  - Null space dim $N-2$
- Eigenvectors form a 2d space
  - sin and cos basis for the circular attractor.

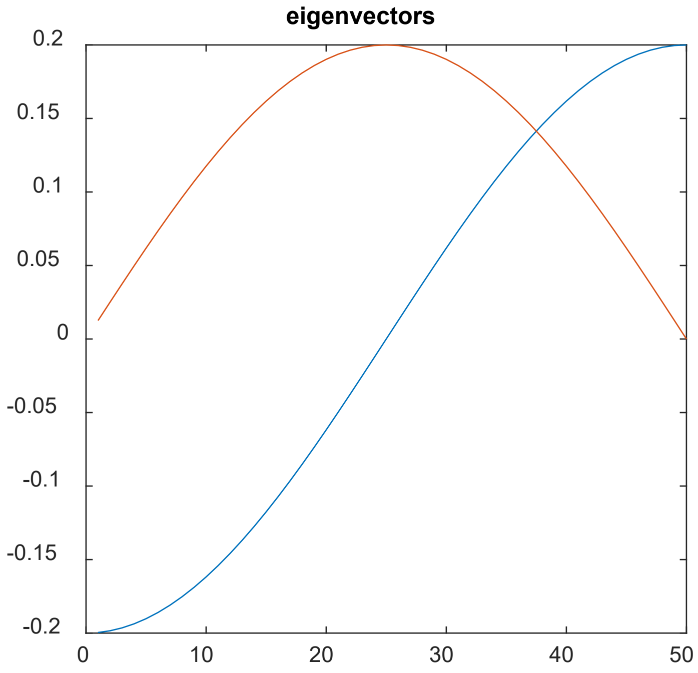


Eigen Structure of a Linear Ring Attractor Model

```
%Matlab simulation code
prefang = 0:pi/50:pi-0.0001;
angdiff = prefang - prefang';
wmat = cos(angdiff);
[evc,eva] = eig(wmat);
```







Neuro120_tutorials/RingAttractorEigs.m at master · Animadversio/Neuro120_tutorials

The only two non-zero eigenvector form a sin cos basis of the ring.