EPFL

Eigenvalues & Eigenvectors EE-312

Prof. Pierre Vandergheynst

Diagonalization - general case

$A \in \mathbb{C}^{n\times n}$ **with distinct eigenvalues** $\lambda_1, \ldots, \lambda_n$

Corresponding right eigenvectors $X = [x_1, \cdots, x_n]$ form a linearly independent set

left eigenvectors $Y = [y_1, \cdots, y_n]$

No orthogonality ? No but "bi-orthogonality"

$Ax_i = \lambda_i x_i$ and $y_j^H A = \lambda_j y_j^H$ $\quad \lambda_i \neq \lambda_j$

Then $y_j^H x_i = 0$

x_i cannot be orthogonal to y_i as well (since they are independent)

So we can normalise the y_j's and/or the x_i's so that $y_i^H x_i = 1, \forall i = 1, \cdots n$

$$A = X\Lambda X^{-1} = X\Lambda Y^H = \sum_{i=1}^{n} \lambda_i x_i y_i^H$$

Diagonalization - special case

Hermitian matrices (symmetric in the real-valued case) : real eigenvalues

All eigenvectors of distinct eigenvalues are orthogonal
Moreover A has n linearly independent eigenvectors and it is always possible to find an orthonormal basis of eigenvectors of A
(even if some eigenvalues are degenerate)

If moreover A is positive definite $x^T A x > 0$, all eigenvalues are positive $\lambda_i > 0$
A is positive semi-definite $x^T A x \geq 0$, all eigenvalues are $\lambda_i \geq 0$

The Spectral Theorem

If $A \in \mathbb{C}^{n \times n}$ is hermitian, there exists a orthonormal basis of $\mathbb{C}^n$ of eigenvectors of A

<u>Main idea of the proof:</u>

By the fundamental theorem of algebra applied to $\pi(\lambda)$, there is an eigenvalue λ_1 and a corresponding eigenvector v_1

Let W_1 be the orthogonal complement to v_1 in $\mathbb{C}^n$

$$\forall w \in W_1 : v_1^H A w = \overline{\lambda_1} v_1^H w = 0 \Rightarrow Aw \in W_1$$

A can be restricted to W_1 and we can proceed by induction

The Spectral Theorem

This decomposition expresses a nice factorization of hermitians matrices

$$X^{-1}AX = \Lambda \text{ or } A = X\Lambda X^{-1}$$

A is unitarily equivalent to the diagonal matrix Λ

where Λ is the diagonal matrix of eigenvalues (with multiplicity)

This also means you can write the linear transformation as a direct sum of orthogonal projections on the eigenspaces:

$$A = \sum_i \lambda_i P_{\lambda_i}$$

The Spectral Theorem

A fundamental reason the spectral theorem is so important :

Let f be an analytic function, i.e $f(t) = \sum_{k=0}^{+\infty} a_k t^k$ ex: sin, exp, ...

$A \in \mathbb{R}^{n \times n}$ such that $A = U \Lambda U^T$

Then $f(A) = U f(\Lambda) U^T$ where $f(\Lambda) = \text{diag}(f(\lambda_1), \cdots, f(\lambda_n))$

Proof: just think of eigendecomposition of A^k

This is crucial when studying differential equations $\frac{d}{dt}x(t) = Ax(t)$

Solving linear homogeneous systems of ODEs

$$\dot{x}(t) = Ax(t),\ x(t_0) = x_0 \text{ with } x, x_0 \in \mathbb{R}^n \text{ and } A \in \mathbb{R}^{n\times n}$$

Define the matrix exponential: $e^A = \sum_{k=0}^{+\infty} \frac{1}{k!} A^k$

Solution ($t \geq t_0$): $x(t) = e^{(t-t_0)A} x_0$

Proof: use the definition of the matrix exponential and differentiate under the sum

Higher Order Systems

$$x^{(k)}(t) = A_{k-1}x^{(k-1)}(t) + \cdots + A_0 x(t) \text{ where } A_i \in \mathbb{R}^{n\times n}$$

$$\text{and } x^{(k)}(t) = \frac{d^k}{dt^k}x(t)$$

Can be written as a larger order 1 ODE:

$$z(t) = \begin{bmatrix} x(t) \\ x^{(1)}(t) \\ \vdots \\ x^{(k-1)}(t) \end{bmatrix} \in \mathbb{R}^{nk} \qquad \dot{z}(t) = \begin{bmatrix} 0 & \mathbb{I} & 0 & \cdots & 0 \\ 0 & 0 & \mathbb{I} & \cdots & 0 \\ \vdots & & & & \vdots \\ A_0 & A_1 & A_2 & \cdots & A_{k-1} \end{bmatrix} z(t)$$

Modal Decomposition

Suppose now A is diagonalisable: $A = X\Lambda Y^H$

The matrix exponential can be written as: $e^A = \sum_{i=1}^{n} e^{\lambda_i} x_i y_i^H$

Therefore the solution of the homogeneous ODE becomes:

$$x(t) = \sum_{i=1}^{n} e^{\lambda_i (t-t_0)} (y_i^H x_0) x_i$$

Modal directions are invariant

Modal velocities

Long-term behavior

$A \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1, \dots, \lambda_n$

Consider the system: $\dot{x}(t) = Ax(t)$

Solutions involve only functions like: $t^k e^{\lambda_j t}$

Therefore the long-term evolution of solutions is governed by eigenvalue, i.e

$$\lim_{t\to+\infty} \|x(t)\| = \begin{cases} 0 & \text{if } \mathrm{Re}(\lambda_j) < 0 \quad \forall j \\ +\infty & \text{if } \mathrm{Re}(\lambda_j) > 0 \quad \text{for some } j \end{cases}$$

Example

Consider a damped pendulum: $mx^{(2)}(t) = -kx(t) - bx^{(1)}(t)$

$$z(t) = \begin{bmatrix} x(t) \\ x^{(1)}(t) \end{bmatrix} \qquad \dot{z}(t) = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix} z(t)$$

You can now solve the system and check how $z(t)$ evolves as a function of the eigenvalues of the system.

Linear Homogeneous Systems of Difference Equations

Discrete-time equivalent to ODEs: $x[k+1] = Ax[k]$; $x[0]$ given

It is easy to see that: $x[k] = A^k x[0] \;\; k \geq 0$

And again, if A is diagonalizable this simplifies thanks to: $A^k = X\Lambda^k Y^H = \sum_{i=1}^{n} \lambda_i^k x_i y_i^H$

Asymptotic Behaviour of Linear Dynamical Systems

The eigendecomposition can also give us insights about the long-term evolution of the system

Example: discrete-time case $x[k+1] = Ax[k];\ x[0]$ given

We already know: $x[k] = A^k x[0] \;\; k \geq 0$

$$x[k] = A^k(c_1 x_1 + \cdots + c_n x_n)$$
$$= c_1 \lambda_1^k x_1 + \cdots + c_n \lambda_n^k x_n$$

If we have $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$

dominant eigenvalue

$$\lim_{k \to +\infty} x[k] \simeq \lambda_1^k c_1 x_1$$

system evolves in the direction of the dominant eigenvector

Inhomogeneous systems

$\dot{x}(t) = Ax(t) + Bu(t); \quad x(t_0) = x_0 \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m, \quad B \in \mathbb{R}^{n\times m}$ given

"Variation of parameters" formula $t \geq t_0$

$$x(t) = e^{(t-t_0)A}x_0 + \int_{t_0}^{t} e^{(t-s)A}Bu(s)ds$$

homogeneous case

At each time, a new solution with initial conditions is computed and added

Matrices and Eigendecompositions of Networks

Networks can be represented by matrices whose eigenvectors and eigenvalues have many interesting applications!

Undirected networks: adjacency matrix

$A[i, j] = 1$ IFF there is an edge linking nodes i and j

A is symmetric and the degree of node i is $d_i = \sum_j A[i, j]$

$D = \text{diag}(d_i)$ is the degree matrix of the network

Matrices and Eigendecompositions of Networks

Undirected networks:

The (combinatorial) Laplacian of the network is $L = D - A$

$L \in \mathbb{R}^{n \times n}$ is symmetric and positive semi-definite, i.e $x^T L x \geqslant 0 \quad \forall x \in \mathbb{R}^n$

It has only real, non-negative eigenvalues and an ortho. basis of eigenvectors

The smallest eigenvalue is $\lambda_1 = 0$ and its multiplicity $= \{\#$ connected components$\}$

If a network is connected but has k clusters or communities there will typically be a gap between λ_k and λ_{k+1}. This is the basis of a famous algorithm...

Matrices and Eigendecompositions of Networks

Spectral Clustering

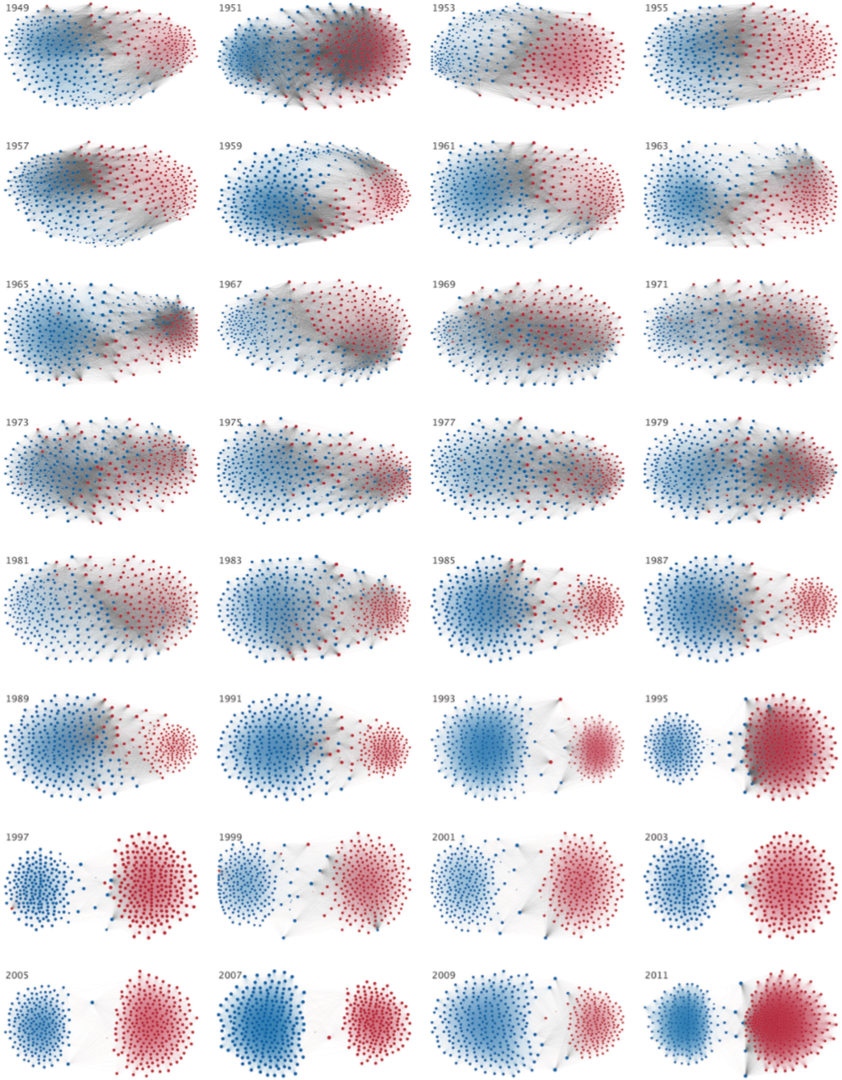
Input: adjacency matrix of graph with k communities or clusters

Compute: eigenvectors x_i corresponding to k smallest eigenvalues

$$X = \begin{pmatrix} | & \dots & | \\ x_1 & \dots & x_k \\ | & \dots & | \end{pmatrix}$$

k-dimensional graph features
will be similar for all nodes in same community

apply clustering to the feature matrix (group similar features)


1949
1951
1953
1955
1957
1959
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011

Matrices and Eigendecompositions of Networks

Networks can be represented by matrices whose eigenvectors and eigenvalues have many interesting applications!

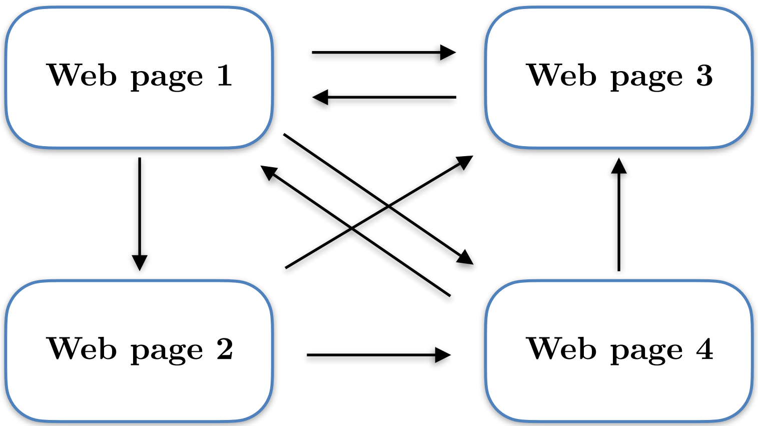
Directed networks: Directed adjacency matrix

$A[i,j] = 1$ IFF there is an edge from i to j

A is not symmetric, nodes have in and out degrees

PageRank:
The $1611 billion eigenvector

Web as a directed graph: nodes = web pages, edges = hyperlinks

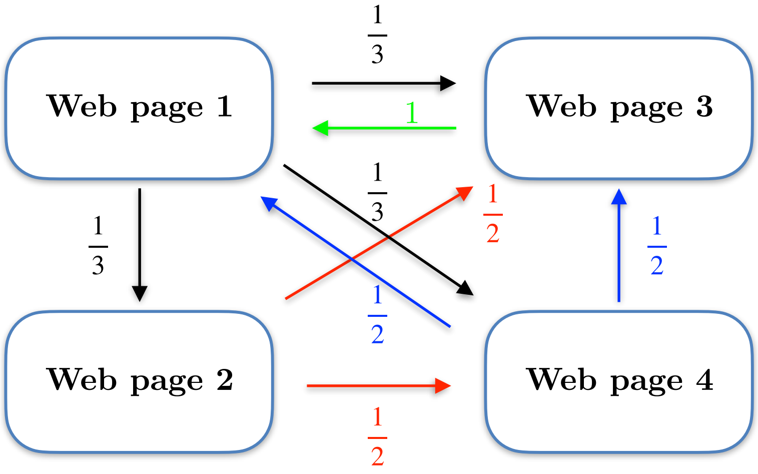


A web page is authoritative if many pages point to it
or if authoritative pages point to it

Challenge: Compute a ranking of web pages by authority

PageRank:
The $1611 billion eigenvector

Hyp. 1: Each page transfers its authority equally to pages it links to



Source page

$$A = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

Target page

Hyp. 2: Each link to a page increases its importance $v \mapsto Av$

PageRank: The \$1611 billion eigenvector

Source page

$$A = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

Target page

Dynamical systems view:

$$v[0] = \frac{1}{\text{numb. pages}}$$

$$v[k] = A^k v[0]$$

All matrices of this form ("column stochastic") have a dominant eigenvalue $= 1$

$$\lim_{k \to +\infty} v[k] = x_{\text{dominant}}$$

In our example: $x_{\text{dominant}} = \begin{pmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{pmatrix}$

Understanding Recurrent Neural Networks (RNNs)

Let's consider a state variable following a first order ODE: $\dot{h}(t) = f(h(t))$

$h(t) \in \mathbb{R}^n$ and $f()$ is a component-wise non-linearity

This can be discretised in time using the forward Euler method:

$$h_n = h_{n-1} + Tf(h_{n-1})$$

which has the form of a Recurrent Neural Net with no input

Understanding Recurrent Neural Networks (RNNs)

Consider the particular case: $\dot{h}(t) = f(Wh(t) + b)$ $W \in \mathbb{R}^{n \times n}$ Weight matrix

We will study how the solution $h(t)$ depends on initial conditions $h(0)$

$$\frac{d}{dt}\left(\frac{\partial h(t)}{\partial h(0)}\right) = J_F(h(t))\frac{\partial h(t)}{\partial h(0)}$$

$F(x) = f(Wx + b)$ Jacobian: $J_F(x)_{ij} = \frac{\partial F_i}{\partial x_j}(x)$

$J_F(x) = \operatorname{diag}\left(f'(Wx + b)\right) W$

Understanding Recurrent Neural Networks (RNNs)

$$\frac{d}{dt}\left(\frac{\partial h(t)}{\partial h(0)}\right) = J_F(h(t))\frac{\partial h(t)}{\partial h(0)}$$

as a function of t and initial conditions

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

$$A(0) = \mathbb{I}_n$$

$$A_{ij}(t) = \frac{\partial h_i(t)}{\partial h_j(0)}$$

If the Jacobian changes slowly with time, we therefore have:

$$A(t) = e^{tJ_F}$$

And we see that the eigenvalues of the Jacobian will play a crucial role!

Aside: stability of initial value problems

Stability (informal definition)**:**

Consider a solution $h(t)$ of the initial value problem $\dot{h}(t) = f(h(t)) \quad h(0) = h_0$

Let $z(t)$ be a solution with $z(0) = z_0$

Stable (Lyapunov stable): $\forall \delta > 0, \exists \epsilon > 0$ s.t

$$\|h_0 - z_0\| \leq \delta \Rightarrow \|h(t) - z(t)\| \leq \epsilon, \forall t$$

Asymptotically stable: $\exists \delta > 0$ s.t

$$\|h_0 - z_0\| \leq \delta \Rightarrow \lim_{t \to +\infty} \|h(t) - z(t)\| = 0$$

Another point of view: perturbation analysis

$$\dot{h}(t) = f(h(t))$$

$$z(t) = h(t) + \epsilon d(t)$$

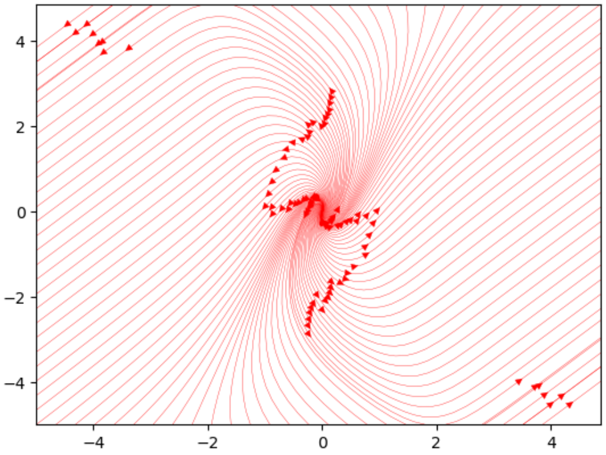
$$\begin{aligned} \dot{z}(t) &= \dot{h}(t) + \epsilon \dot{d}(t) \\ &= f\big((h(t) + \epsilon d(t)\big) \\ &= f\big((h(t))\big) + \epsilon J_f d(t) \end{aligned}$$

$$\dot{d}(t) = J_f d(t)$$

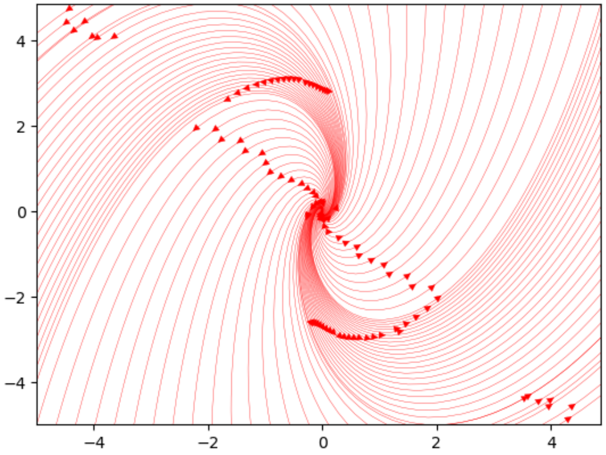
Again the spectrum of the Jacobian will inform us whether perturbations can grow, decay or sustain.

Aside: stability of initial value problems

Original RNN

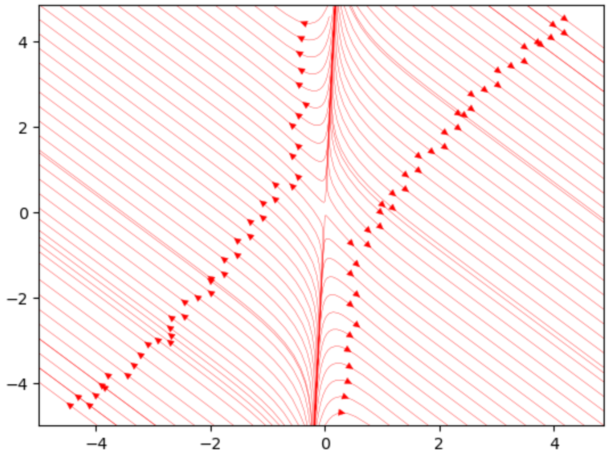


Dynamics linearized around (0,0)

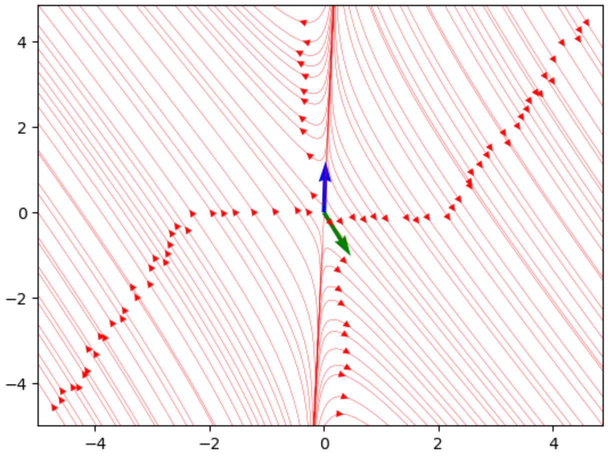



```
W=[ 0.64952629, -0.22683761],
        [ 1.40245917,  0.55148058]]
```

Original RNN

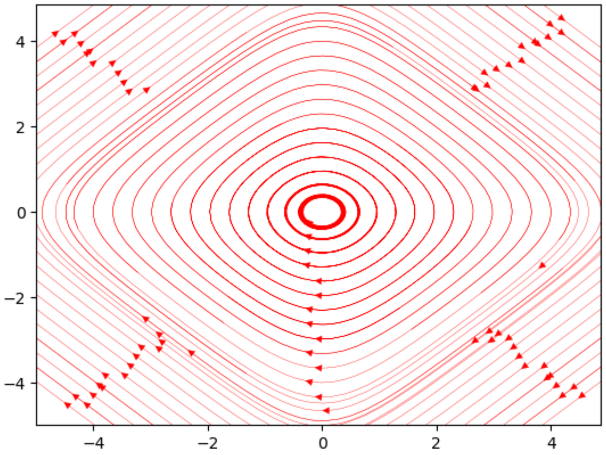


Dynamics linearized around (0,0)

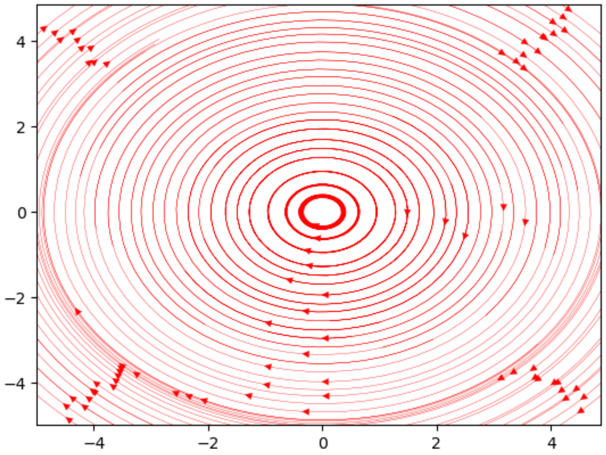



```
W= array([[ 1.20163856, -0.05038193],
          [-2.33322499, -0.14888851]])
```

Original RNN



Dynamics linearized around (0,0)




```
W= array([[ 0, 1],
          [-1, 0]])
```

Back to the discretization !

Question: when is Euler's approximation a stable discrete system ?

$$h_n = h_{n-1} + Tf(h_{n-1})$$

Again we know that with linearization/perturbation analysis, we can just look at how this will behave in the simplest linear case and in the Jacobian eigen-basis :

$$u_n = u_{n-1} + T\lambda u_{n-1}$$

$$u_{n+1} = (1 + T\lambda)u_n \rightarrow u_n = (1 + T\lambda)^n u_0$$

We therefore need: $|1 + T\lambda| < 1$