Chapter 9

Dynamical Systems Review

Neural Networks: A Classroom Approach
Satish Kumar
Department of Physics & Computer Science
Dayalbagh Educational Institute (Deemed University)

Global Behaviour of Systems

- Global behaviour of systems is a manifestation of local interactions between components
- Generally characterized by a set of variables called *states*
- Example: inductor current and capacitor voltage in an electrical circuit

States, State Vectors

- *State variables* of the system characterize its behaviour
- A vector of states is called a *state vector*, $X \in \Re^n$
- Example: the vector of neuronal activations is a state vector of the neuron layer

Feedforward Systems

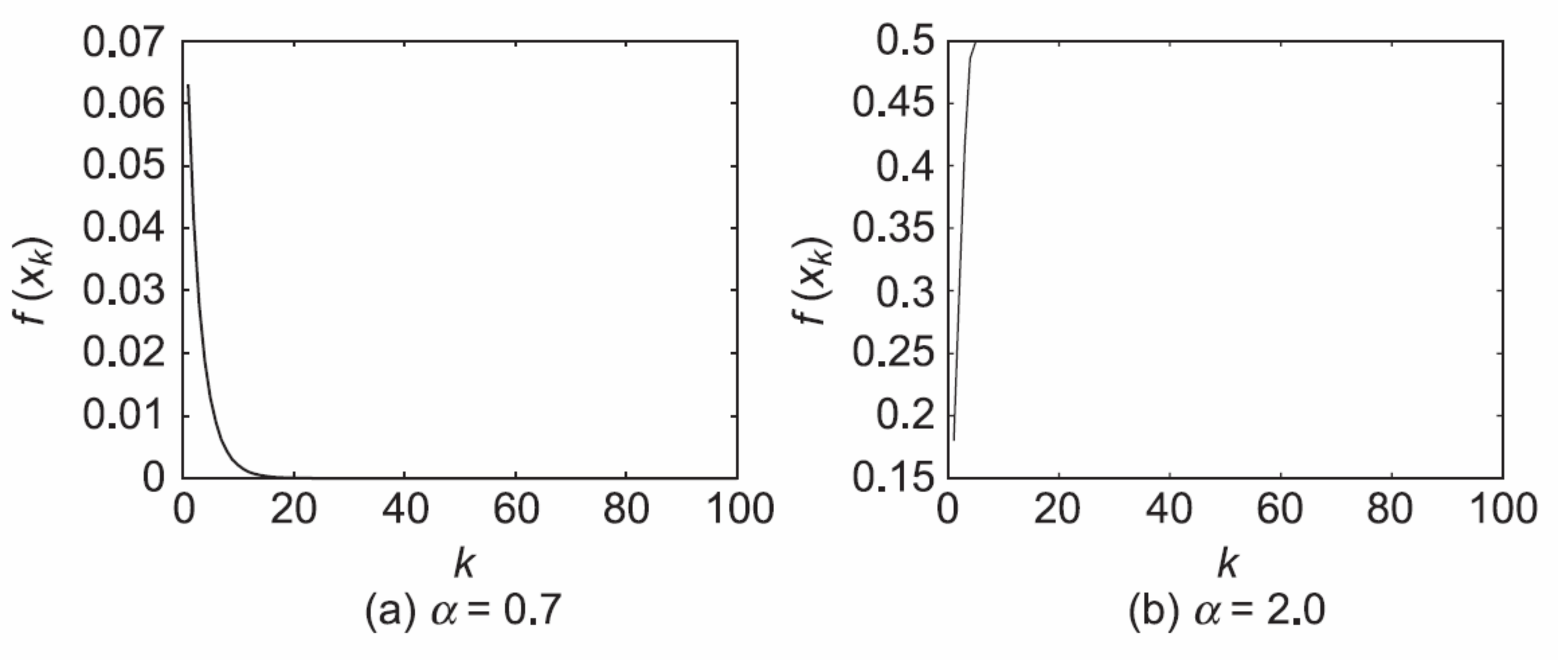
- Inputs set up internal states and generates an output which remains unchanged as long as the input is held constant
  - Combinational logic circuit
  - Feedforward neural network

Feedback System

- Output fed back as an additional input to the system after delay modifies the overall input to the system
- Changes the internal "state" of the system which in turn generates a new set of outputs
- States of the system evolve in time
  - *Time dynamical system*

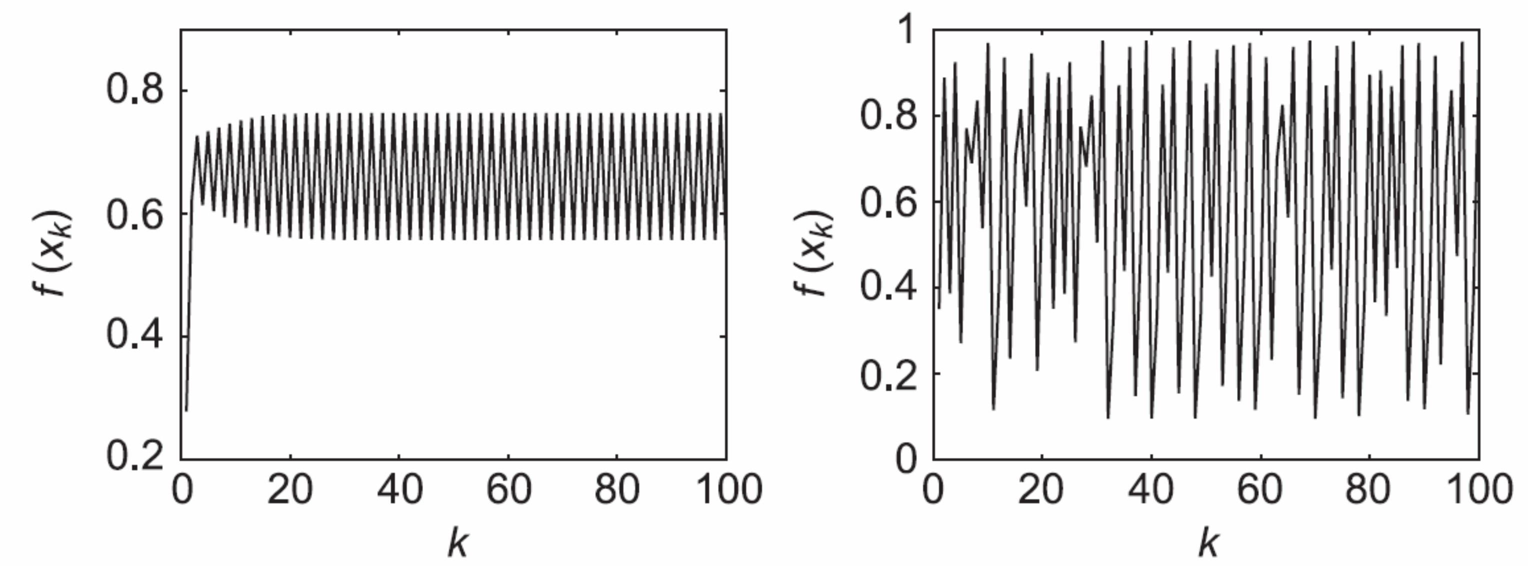
Discrete Time Logistic Function

- $f(x) = a\,x\,(1 - x),\ a \in [0, 4]\ ,\ x \in (0, 1)$
  - Difference form: $x_{k+1} = a\,x_k\,(1 - x_k)$



Discrete Time Logistic Function

- The logistic function undergoes bifurcations from fixed point behaviour, through limit cycles towards chaos as the gain is changed from 0 towards 4



(c) $\alpha = 3.1$ (d) $\alpha = 3.9$

State Equations

- An n-dimensional system is governed by the equation

$$\dot{X} = f(X) \qquad X(0) = X_0$$

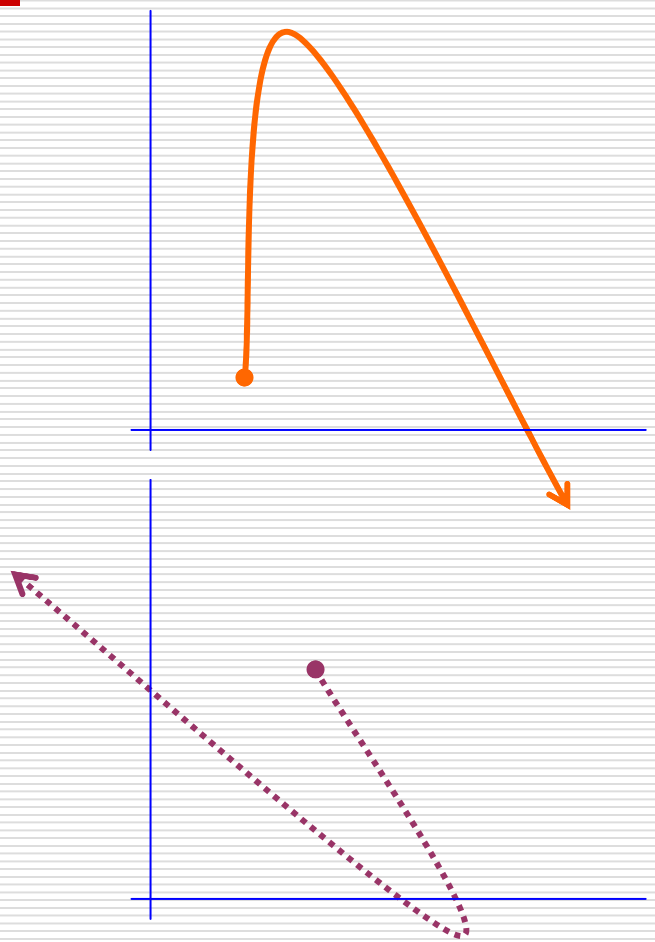
Vector field

- Autonomous system: vector field not a function of time
- Non-autonomous system: vector field is a function of time

$$\dot{X} = f(X, t) \qquad X(t_0) = X_0$$

Trajectories and Orbits

- Time is continuous
  - $t \in \Re$
  - Continuous time dynamical system
  - Trajectory
- Time is discrete
  - $t \in Z^+$
  - Discrete time dynamical system
  - Orbit
- Collection of trajectories is called a phase portrait



Existence and Uniqueness Theorem

- Consider the initial value problem

$$\dot{X} = f(X), \;\; X(0) = X_0$$

- Suppose that $f(.)$ is continuous, and that all its partial derivatives $\partial f_i/\partial x_j$, $i,j = 1,\ldots,n$, are continuous on some open connected set $D \subset \Re^n$. Then for $X_0 \in D$, the initial value problem has some solution $X(t)$ on some interval $(-\tau, \tau)$ about $t = 0$, and the solution is unique.

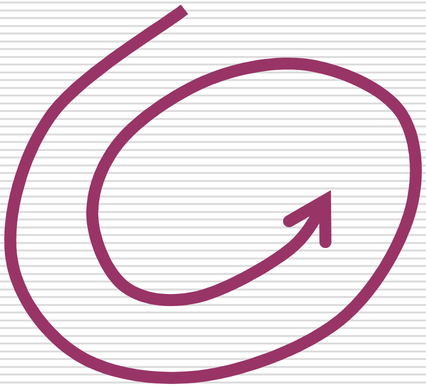
Attractors

- Attractors: regions of state space to which neighbouring trajectories converge
- Take on different shapes and sizes depending on the system under consideration
- Examples:
  - Fixed points
  - Limit cycles

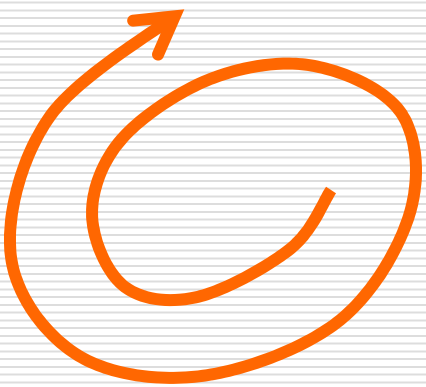
Definition

- Attractor, A
  - A closed set satisfying the following properties
    - Invariant set – trajectories starting within A remain in A
    - A attracts an open set of initial conditions called the basin of attraction
    - No proper subset of A satisfies the above properties

Attractors and Repellers



Trajectories generally settle down to some part of the state space



Trajectories do not settle down and eventually shoot off to infinity

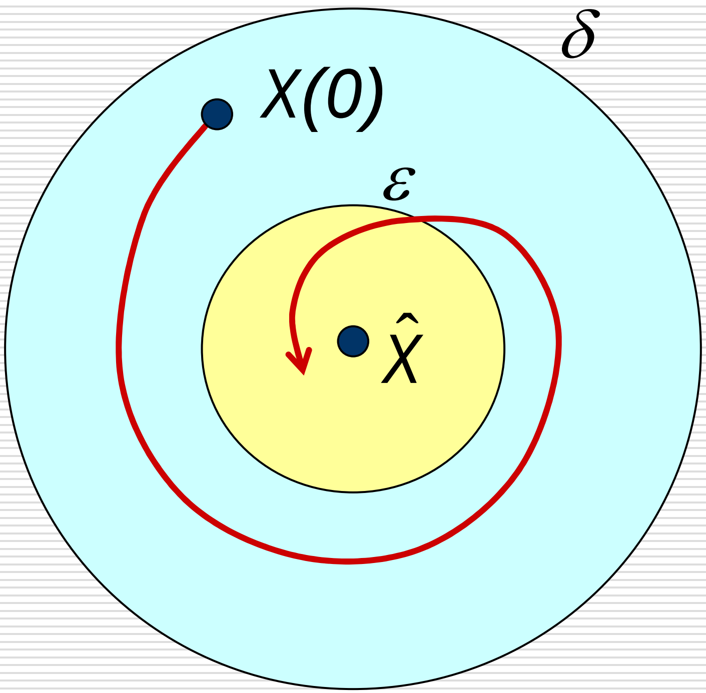
Stability

- Uniform stability
  - $$\|X(0)-\hat{X}\| \;<\; \delta \implies \|X(t)-\hat{X}\| \;<\; \epsilon, \quad \forall\, t > 0$$
  - Positive ε, δ
- Convergent

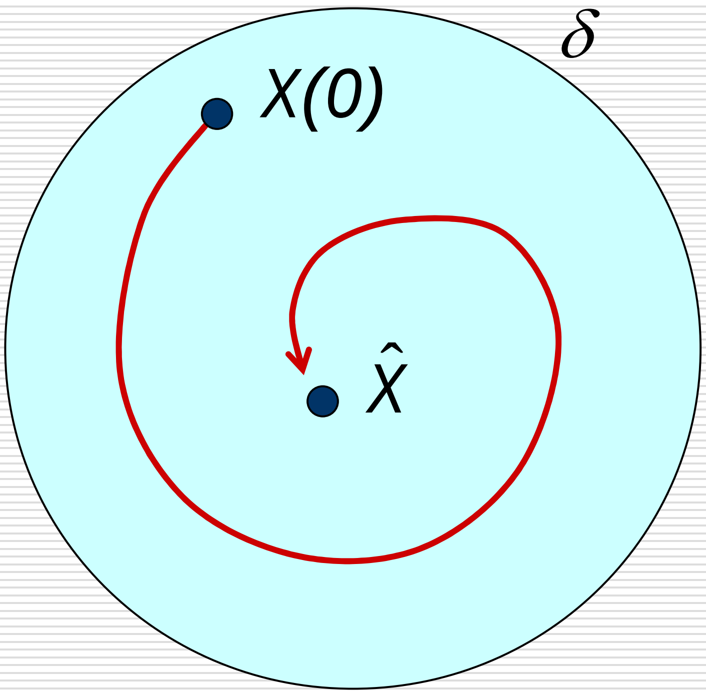
$$\|X(0)-\hat{X}\| \;<\; \delta \quad \implies \quad X(t) \to \hat{X} \quad \text{as} \quad t \to \infty$$

  - Positive δ
- Asymptotically stable
  - Uniformly stable and convergent

Uniform and Convergent Stability



Uniform stability



Convergent stability

Asymptotic Stability

- An equilibrium state X is asymptotically stable in the large or globally asymptotically stable if it is stable and all trajectories converge to X as $t \to \infty$

Autonomous Linear Systems

- Commonly characterized by differential equations of the following form

$$
\begin{aligned}
\dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
\dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
\vdots \quad & \qquad\quad \vdots \qquad\qquad\quad \vdots \\
\dot{x}_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n
\end{aligned}
$$

- More conveniently as $\dot{X} = \mathbf{A}X$

Vector Differential Equation

$$\dot{X} = AX$$

- Interpret A as a linear map
  - $A: \Re^n \rightarrow \Re^n$
- An equilibrium state of the system is the point in space where the vector field vanishes
- For the autonomous case, the origin is a fixed point

Matrix Diagonalization

- Assume matrix A is non-singular
- A has n distinct eigenvalues
- Orthogonal similarity transformation
  - $\dot{X} = AX$ transforms to $\dot{Y} = DY$
  - $D = P^{-1}AP$, P is a matrix with eigenvectors as columns

Eigensolution of Linear Systems

- Solution of $\dot{X} = AX$
- In terms of eigenvectors and eigenvalues

$$X(t) = c_1 e^{\lambda_1 t} \eta_1 + c_2 e^{\lambda_2 t} \eta_2 + \cdots + c_n e^{\lambda_n t} \eta_n$$

Two Dimensional Linear System

- Helps develop and intuitive understanding

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

- Eigenvalues
- Whether they are real or complex depends on $\xi - 4\Delta^2$

$$\lambda_1 = \frac{\zeta + \sqrt{\zeta^2 - 4\Delta}}{2}$$

$$\lambda_2 = \frac{\zeta - \sqrt{\zeta^2 - 4\Delta}}{2}$$

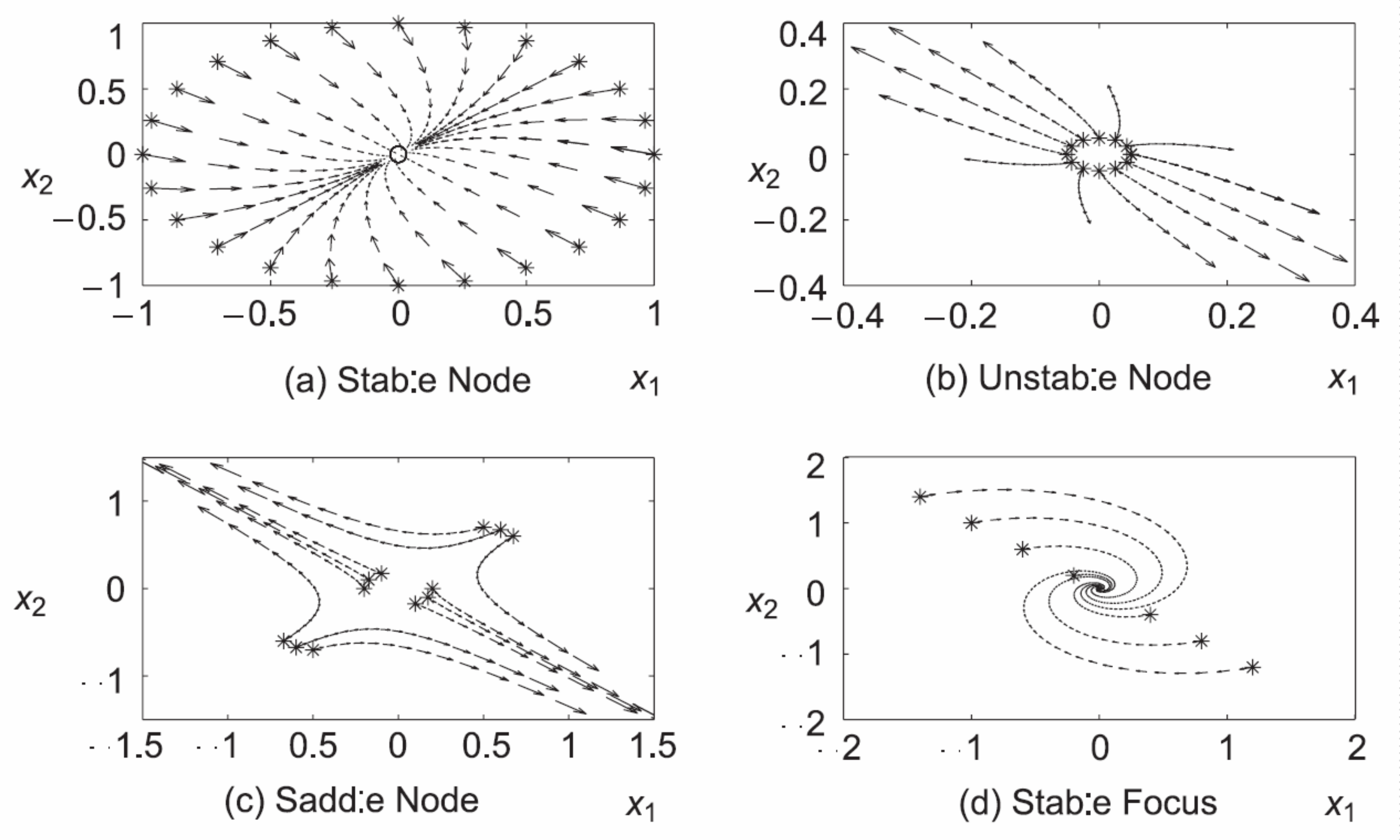
trace of A

determinant of A

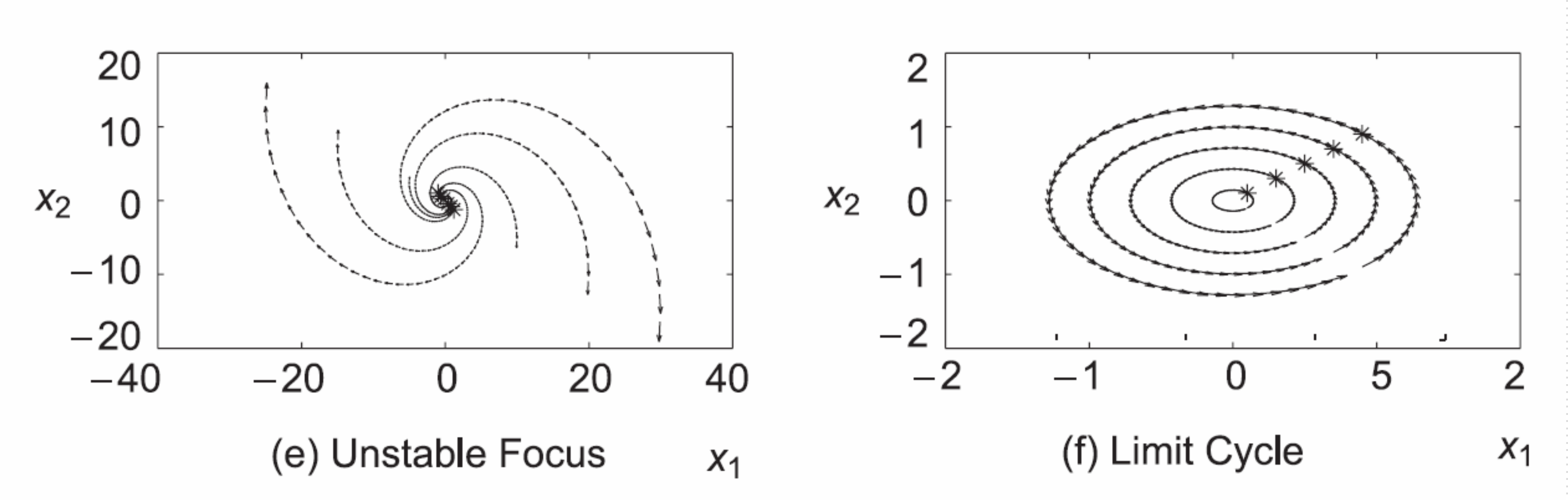
Two Dimensional Linear System: Behavioural Patterns

- With real eigenvalues $\xi - 4\Delta^2 > 0$ and the solution is either a
  - stable node $\lambda_1 < 0, \ \lambda_2 < 0$
  - unstable node $\lambda_1 > 0, \ \lambda_2 > 0$
  - saddle point $\lambda_1 < 0, \ \lambda_2 > 0$
- With complex eigenvalues $\xi - 4\Delta^2 < 0$ and the eigenvalues take the form $\alpha \pm j\beta$. The solution then takes the form
  - stable focus $\alpha < 0$
  - unstable focus $\alpha > 0$
  - limit cycle or center $\alpha = 0$

Two Dimensional Linear System: Behavioural Patterns



Two Dimensional Linear System: Behavioural Patterns



Summary of the Behaviour of Linear Systems

Eigenvalues of Matrix A	Type of equilibrium state $\hat{X}$
Real and negative	Stable node
Complex conjugate with negative real parts	Stable focus
Real and positive	Unstable node
Complex conjugate with positive real parts	Unstable focus
Real with opposite signs	Saddle point
Conjugate purely imaginary	Limit cycle or center

- Linear system: exactly one of the above six kinds of behaviour in the *entire* state space.
- Need not hold for non-linear systems

Non-linear Dynamical Systems

- Autonomous non-linear systems can be described by the vector differential system

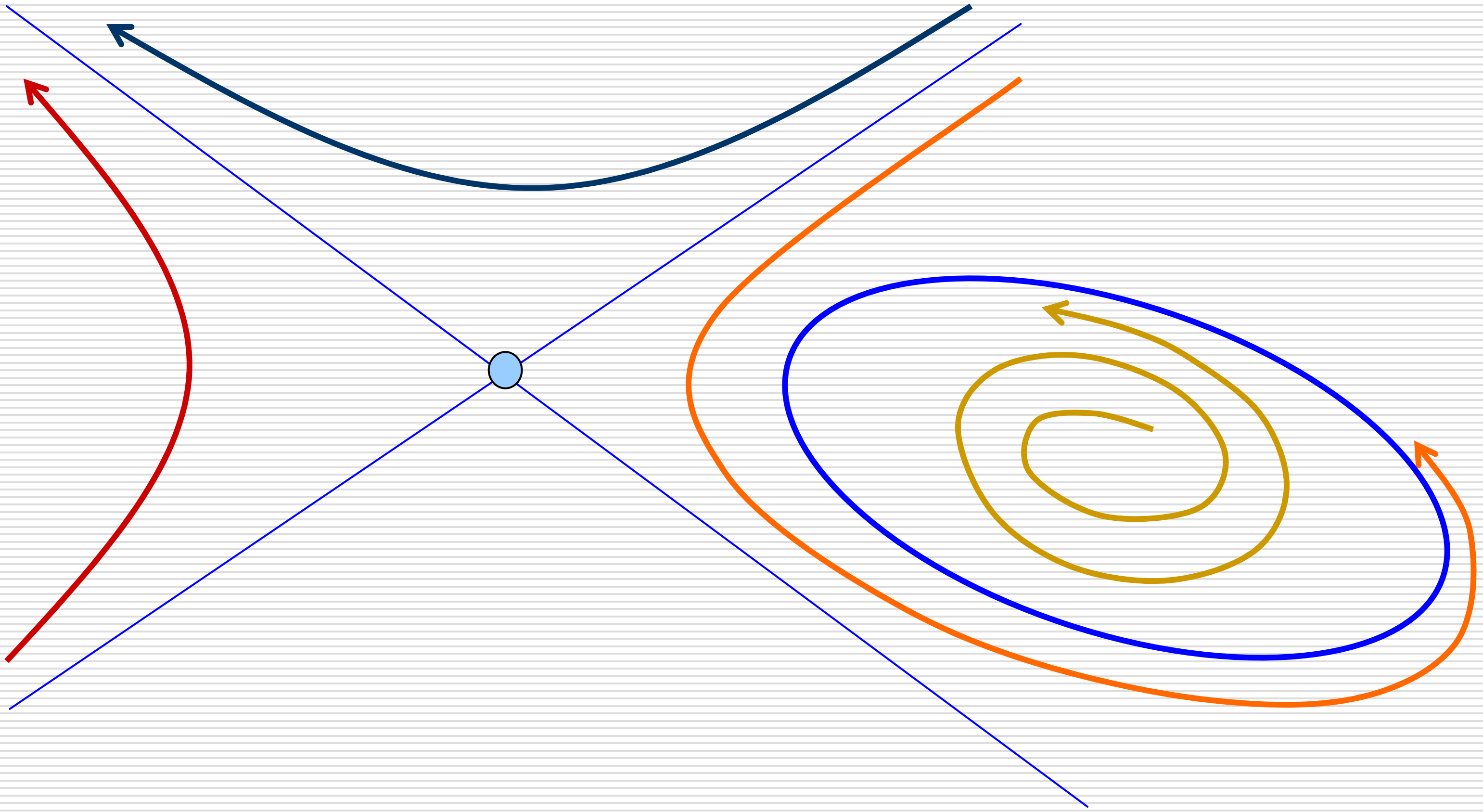
$$\dot{X} = f(X)$$

Vector field f is non-linear

Non-linear Systems: Difference from Linear Systems

- Presence of multiple attractors
- Structure of attractors is often a sensitive function of system parameters
  - Change in the structure of attractors when a specific system parameter changes is called a *bifurcation*
  - Example: Attracting fixed points can suddenly become repellers!

Attractors of Non-linear Systems can be of Multiple Kinds



Linearization of Non-Linear Systems

- Use the 2-dimensional case as an example

$$\dot{x}_1 = f_1(x_1, x_2)$$

$$\dot{x}_2 = f_2(x_1, x_2)$$

Linearization of Non-Linear Systems

- Use a Taylor series expansion

$$\dot{x}_1 = f_1(\hat{x}_1, \hat{x}_2) + a_{11}(x_1 - \hat{x}_1) + a_{12}(x_2 - \hat{x}_2) + \mathcal{O}_1(\cdot)$$

$$\dot{x}_2 = f_2(\hat{x}_1, \hat{x}_2) + a_{21}(x_1 - \hat{x}_1) + a_{22}(x_2 - \hat{x}_2) + \mathcal{O}_2(\cdot)$$

- where

$$a_{11} \triangleq \left. \frac{\partial f_1(x_1, x_2)}{\partial x_1} \right|_{x_1=\hat{x}_1, x_2=\hat{x}_2} \qquad a_{12} \triangleq \left. \frac{\partial f_1(x_1, x_2)}{\partial x_2} \right|_{x_1=\hat{x}_1, x_2=\hat{x}_2}$$

$$a_{21} \triangleq \left. \frac{\partial f_2(x_1, x_2)}{\partial x_1} \right|_{x_1=\hat{x}_1, x_2=\hat{x}_2} \qquad a_{22} \triangleq \left. \frac{\partial f_2(x_1, x_2)}{\partial x_2} \right|_{x_1=\hat{x}_1, x_2=\hat{x}_2}$$

Linearization of Non-Linear Systems

- Re-write in the following form noting that $\frac{d}{dt}(x_1 - \hat{x}_1) = \dot{x}_1$

$$\frac{d}{dt}(x_1 - \hat{x}_1) = a_{11}(x_1 - \hat{x}_1) + a_{12}(x_2 - \hat{x}_2) + \mathcal{O}_1(\cdot)$$
$$\frac{d}{dt}(x_2 - \hat{x}_2) = a_{21}(x_1 - \hat{x}_1) + a_{22}(x_2 - \hat{x}_2) + \mathcal{O}_2(\cdot)$$

- Define $\tilde{x}_1 \triangleq (x_1 - \hat{x}_1)$

$$\tilde{x}_2 \triangleq (x_2 - \hat{x}_2)$$

Linearization of Non-Linear Systems

- Finally yields,

$$\frac{d}{dt}\begin{pmatrix}\tilde{x}_1\\ \tilde{x}_2\end{pmatrix} = \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix}\tilde{x}_1\\ \tilde{x}_2\end{pmatrix} + \mathcal{O}(\cdot)$$

- or $\dot{\tilde{X}} = \mathbf{A}\tilde{X} + \mathcal{O}(\cdot)$

$$\mathbf{A} = \left.\begin{pmatrix}\dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2}\\[2ex] \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2}\end{pmatrix}\right|_{X=\hat{X}}$$

Jacobian Matrix

Hartman-Grobman Theorem

- Let $f_1(x_1,x_2)$ and $f_2(x_1,x_2)$ have continuous first order partial derivatives in a neighbourhood of the equilibrium state X_E. Then if the origin of the linearized state equation is a stable (unstable) node, a stable (unstable) focus, or a saddle point, then the trajectories in a small neighbourhood of X_E of the corresponding non-linear state equation will also behave "as" a stable (unstable) node, stable (unstable) focus, or a saddle point respectively.

Analysis of a Non-linear Differential System Through Linearization

- Example:
- Equilibrium point is the origin

$$\dot{x}_1 = -x_2 + a x_1 (x_1^2 + x_2^2)$$

$$\dot{x}_2 = x_1 + a x_2 (x_1^2 + x_2^2)$$

- Jacobian matrix

$$\mathbf{A} = \begin{pmatrix} 3ax_1^2 + ax_2^2 & -1 + 2ax_1x_2 \\ 1 + 2ax_1x_2 & ax_1^2 + 3ax_2^2 \end{pmatrix}$$

$$\mathbf{A}\Big|_{X=\hat{X}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

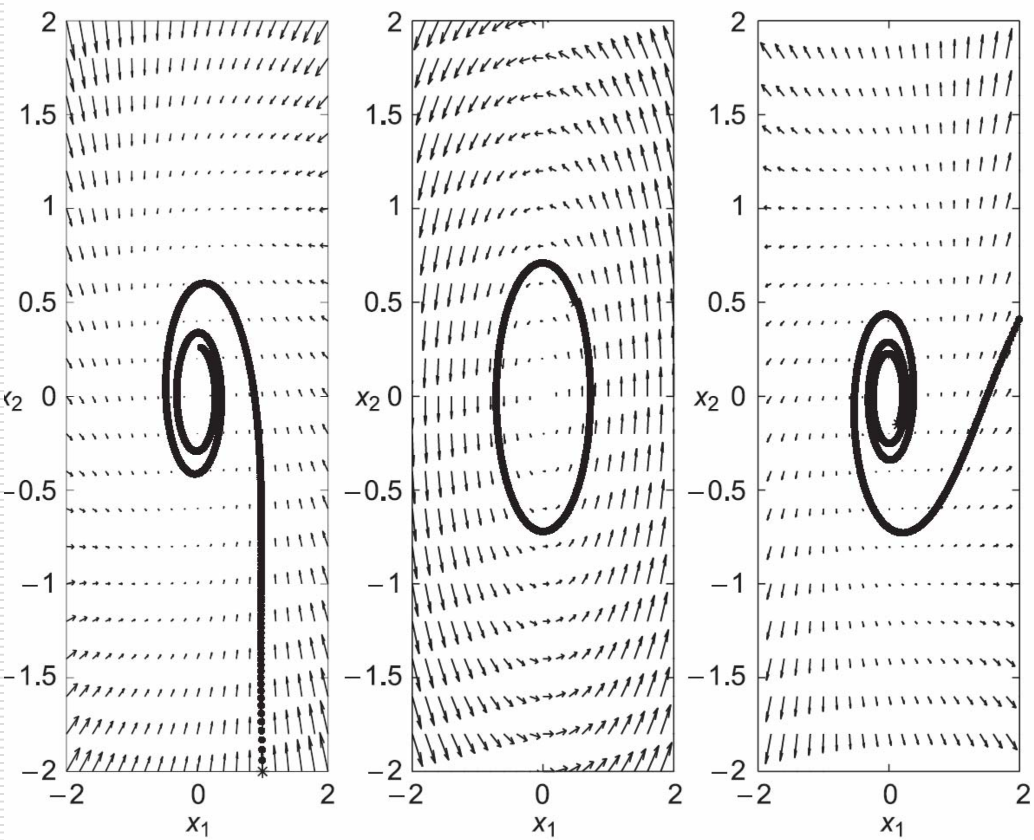
- $\xi = 0$, $\Delta = 1 > 0$
- Linearization predicts origin is a limit cycle (center)
- Prediction is incorrect

Transformation of Variables Yields a Different Solution!

- Polar version of the system is

$$\dot{r} = ar^3$$
$$\dot{\theta} = 1$$

- All trajectories rotate about the origin at a constant angular velocity
  - a = 0: center
  - a < 0: stable focus
  - a > 0: unstable focus



(a) $a = -0.5$ (b) $a = 0$ (c) $a = 0.5$

Lyapunov Stability

- Allows investigation of the stability problem
- Makes use of a continuous scalar function of the state vector, called a *Lyapunov function*
- Straightforward to determine the stability by analyzing the behaviour of this auxiliary function
- Lyapunov's Theorem

Lyapunov Stability Theorem

- Consider a non-linear system $\dot{X} = f(X)$
- $f: W \rightarrow \Re^n$ is a C^1 map on an open set $W \subset \Re^n$.
- Let $\hat{X}$ be an equilibrium point of the system.
- Further, let $E(X): U \rightarrow \Re$ be a continuous scalar function defined on a neighbourhood $U \subset W$ of $\hat{X}$ differentiable on $U \backslash \{\hat{X}\}$
- Then $\hat{X}$ is stable if $E(\hat{X}) = 0$, *and* $E(\hat{X}) > 0$ *if* $X \neq \hat{X}$
  $\dot{E}(X) \leq 0$, *in* $U \backslash \{\hat{X}\}$
- $\hat{X}$ is asymptotically stable if $\dot{E}(X) < 0$, *in* $U \backslash \{\hat{X}\}$

Example of Lyapunov Approach

- Return to our earlier non-linear system example with a = -1

$$\dot{x}_1 = -x_2 - x_1(x_1^2 + x_2^2)$$
$$\dot{x}_2 = x_1 - x_2(x_1^2 + x_2^2)$$

- Define a p.d. scalar function
  - $E(X) = x_1^2 + x_2^2$
- Then,

$$\begin{aligned}\dot{E}(X) &= 2x_1\dot{x}_1 + 2x_2\dot{x}_2 \\ &= 2x_1(-x_2 - x_1(x_1^2 + x_2^2)) + 2x_2(x_1 - x_2(x_1^2 + x_2^2)) \\ &= -2(x_1^2 + x_2^2)^2 \\ &< 0, \quad X \neq (0,0)\end{aligned}$$

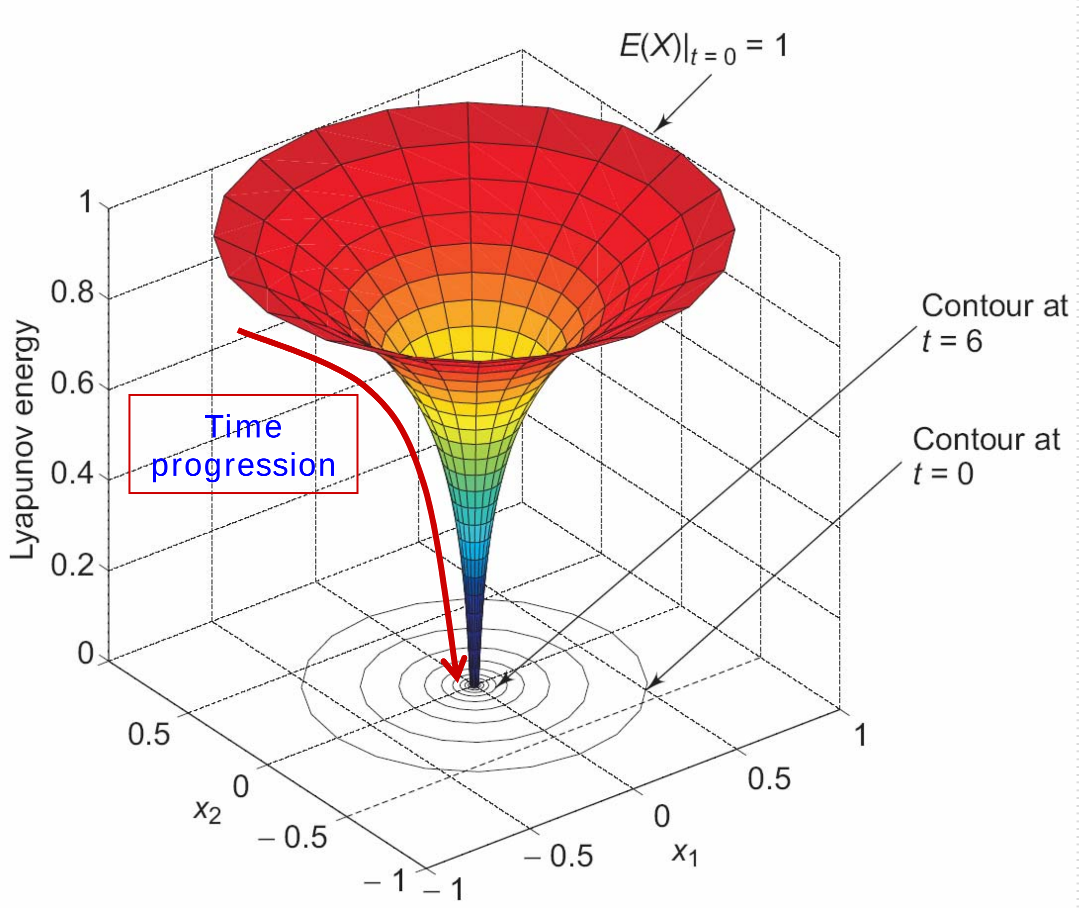
$\Rightarrow$ Global asymptotic stability

Visualization of the Lyapunov's Technique

- Consider the system

$$\dot{x}_1 = -x_1$$
$$\dot{x}_2 = -x_2$$

- Positive definite scalar function
  - $E(X) = x_1^2 + x_2^2$
- Constant energy contours are concentric circles



Gradient Systems

- Sometimes neural network models are gradient systems
- They satisfy the condition,

$$\dot{x}_i = -\frac{\partial E(X)}{\partial x_i}, \qquad i = 1 \ldots n$$

- Stability guaranteed

$$\dot{E}(X) = \sum_{i=1}^{n} \frac{\partial E(X)}{\partial x_i} \dot{x}_i$$

$$= \sum_{i=1}^{n} \dot{x}_i^2$$

$$< 0$$

Quadratic Forms

- Common practice to consider quadratic Lyapunov functions
  - $E(X) = X^T A X$
  - These are positive definite

$$
\begin{aligned}
E(X) &= X^T \left(\sum_{k=1}^{Q} X_k X_k^T \right) X \\
&= \sum_{k=1}^{Q} X^T X_k X_k^T X \\
&= \sum_{k=1}^{Q} \left(X^T X_k \right)^2 \\
&\geq 0
\end{aligned}
$$

Importance for Neural Networks

- Proving the stability of neurodynamical systems
- Sufficient to find a function of network states that is
  - bounded below
  - whose time derivative is negative
- Function value decreases with time, and must eventually hit a lower bound
- Network states cease to evolve any further since the Lyapunov energy can decrease no further.
- Stopping point is to be interpreted as a memory that needs to be recalled

Neurons and Update Strategies

- Common neuron signal functions
  - binary threshold signal function
  - linear threshold signal function
  - sigmoidal signal function.
- Neuron update strategies are
  - *asynchronous update*
    - where neurons update one by one in a purely random order
  - *periodic update*
    - where neurons update one by one in a fixed order in a periodic fashion
  - *parallel or synchronous update*
    - where neurons update together.

Lyapunov Energy Functionals for Neural Networks

- binary threshold logic neurons operating under asynchronous update

$$E(S) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} s_i s_j - \sum_{i=1}^{n} w_{0i} s_i$$

- binary threshold logic neurons operating under parallel update

$$E(S,k) = -\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} s_i^k s_j^{k-1} - \sum_{i=1}^{n} w_{0i}(s_i^k + s_i^{k-1})$$

- sigmoids or linear neurons operating under asynchronous or periodic update

$$E(S) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} s_i s_j + \sum_{i=1}^{n}\int_0^{s_i} \mathcal{S}^{-1}(\alpha)d\alpha - \sum_{i=1}^{n} w_{0i} s_i$$

Lyapunov Energy Functionals for Neural Networks

- Sigmoids or linear neurons operating under parallel update

$$E(S,k) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} s_i^k s_j^{k-1} + \sum_{i=1}^{n}\left[\int_0^{s_i^k} \mathcal{S}^{-1}(\alpha)d\alpha + \int_0^{s_i^{k-1}} \mathcal{S}^{-1}(\alpha)d\alpha\right]$$

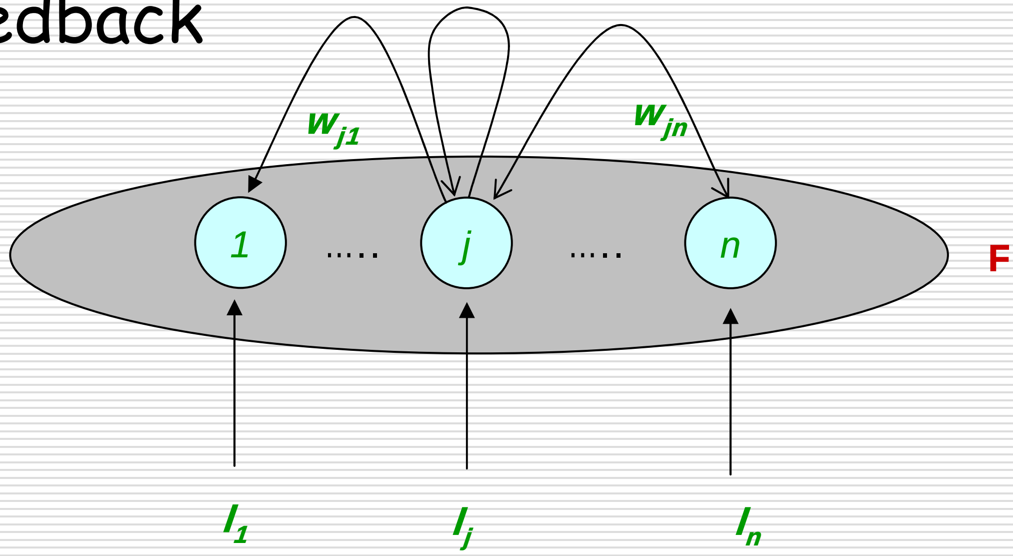
$$-\sum_{i=1}^{n} w_{0i}(s_i^k + s_i^{k-1})$$

- Both the signal and neuron transitions are smooth

$$E(S) = -\sum_{i=1}^{n}\int_0^{s_i} b_i(\alpha_i)\mathcal{S}_i'(\alpha_i)d\alpha_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}s_i s_j$$

Neurodynamical Systems

- High dimensional non-linear systems
  - High dimension: large number of neurons;
  - Non-linearity: the signal function
  - Dynamics: feedback



Neuron Activations

- In a constant state of flux due to three factors:
  - passive decay of neuronal activity;
  - signal feedback from neurons within the same field or from neurons in other fields;
  - external inputs which may excite or inhibit the neuron.

Passive Decay Model

- Assume that the neuronal activation decays in accordance with the first order linear passive decay model:

$$\dot{x}_i = -a_i x_i, \qquad i = 1, \ldots, n$$

- Admits the solution

$$x_i(t) = x(0)\exp(-a_i t), \quad i = 1, \ldots, n$$

- Each of the *n* differential equations is de-coupled due to absence of feedback
- All activations decay smoothly to zero

Adding External Inputs

- Add external inputs to each individual neuron

$$\dot{x}_i = -a_i x_i + I_i, \qquad i = 1, \ldots, n$$

- Admits the solution

$$x_i(t) = x(0)\, exp(-a_i t) + \frac{I_i}{a_i}(1 - exp(-a_i t), \quad i = 1, \ldots, n$$

- Assume that the external input changes much more slowly than the activation
- Initial activity prior to application of the input decays smoothly to zero.

Stability Easily Determined

- Assume a quadratic Lyapunov function

$$E(X) = X^T \mathbf{I} X = \sum_i x_i^2$$

- Substitute the original differential equation into the system

$$\dot{E}(X) = 2\sum_i x_i \dot{x}_i = -2\sum_i a_i (x_i)^2$$

- Negative definite!

$$< 0$$

Additive Neuronal Dynamics

- Add neuronal signal feedback from other neurons in the layer

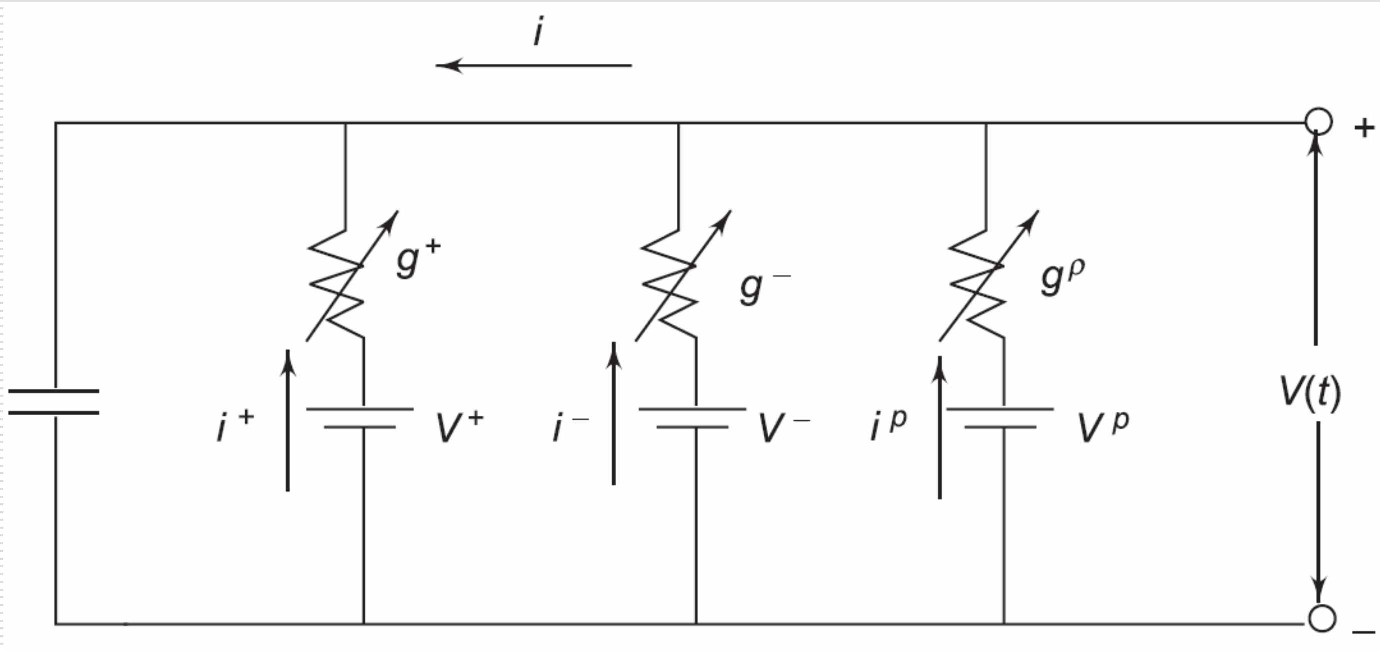
$$\dot{x}_i = -a_i x_i + \sum_{j=1}^{n} w_{ji} S(x_j) + I_i, \qquad i = 1, \ldots, n$$

- Cross-coupled system of differential equations which is non-linear if $S(.)$ is non-linear.

Shunting Neuron Dynamics

- Uses a product of activations and external inputs in place of simple additions used in additive dynamics.
- Embodies the fundamental modelling methodology of the Hodgkin-Huxley model to be introduced in Chapter 13.

Circuit Model of a Neuronal Cell Membrane: Hodgkin-Huxley Equation



$$C\frac{\partial V}{\partial t} = (V^+ - V)g^+ + (V^- - V)g^- + (V^p - V)g^p$$

Maximum (Na)

Minimum (K)

Equilibrium (Cl)

Shunting Neuron Dynamics

- Use the substitutions shown on the right

$$
\begin{aligned}
V(t) &= x_i \\
V^+ &= B_i \\
V^- &= -D_i \\
V^p &= 0 \\
g^+ &= I_i + \textstyle\sum_{j \neq i} s_j w_{ji}^+ \\
g^- &= J_i + \textstyle\sum_{l \neq i} s_l w_{li}^- \\
g^p &= A_i \\
C &= 1
\end{aligned}
$$

$$
\dot{x}_i = -A_i x_i + (B_i - x_i)\left(\sum_{j \neq i} w_{ji}^+ S_j + I_i\right) - (x_i + D_i)\left(\sum_{l \neq i} w_{li}^- S_l + J_i\right), \qquad i = 1, \ldots, n
$$

The Cohen-Grossberg Theorem

- Far reaching implications for neural network theory
- Describes a generalized model of a non-linear dynamical system
- Proves its global asymptotic stability by suggesting an appropriate Lyapunov function.
- Later shown (1989) that a number of major neural network models were indeed special cases of this general system

The Cohen-Grossberg Theorem

- Models that can be written in the form

$$\frac{dx_i}{dt} = a_i(x_i)\left(b_i(x_i) - \sum_{j=1}^{n} c_{ji} d_j(x_j)\right) \quad i = 1, \ldots, n$$

- admit the global Lyapunov function

$$E = -\sum_{i=1}^{n} \int_0^{x_i} b_i(\alpha_i) d_i'(\alpha_i) d\alpha_i + \frac{1}{2} \sum_{j,k=1}^{n} c_{jk} d_j d_k$$

- If the matrix C and functions a_i, b_i, d_j satisfy
  - Symmetry: $c_{ij} = c_{ji}$
  - Positivity: $a_i(x_i) \geq 0$
  - Monotonocity: $d_j(x_j)' \geq 0$

  All trajectories are guaranteed to approach one of possibly infinitely many equilibrium points. (See text for Proof)