Lecture 20

Hopfield Network

- Symmetric weights: $w_{ij} = w_{ji}$
- No self-action: $w_{ii} = 0$
- Zero threshold: $\theta = 0$
- Bipolar states: $s_i \in \{-1, +1\}$
- Discontinuous bipolar activation function:

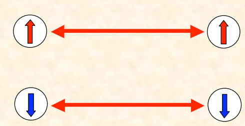
$$\sigma(h) = \operatorname{sgn}(h) = \begin{cases} -1, & h < 0 \\ +1, & h > 0 \end{cases}$$

What to do about $h = 0$?

- There are several options:
  - $\sigma(0) = +1$
  - $\sigma(0) = -1$
  - $\sigma(0) = -1$ or $+1$ with equal probability
  - $h_i = 0 \Rightarrow$ no state change ($s_i' = s_i$)
- Not much difference, but be consistent
- Last option is slightly preferable, since symmetric

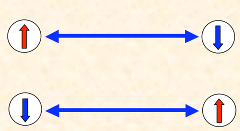
Positive Coupling

- Positive *sense* (sign)
- Large *strength*



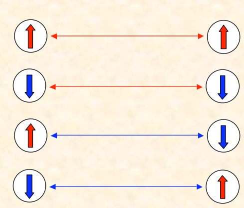
Negative Coupling

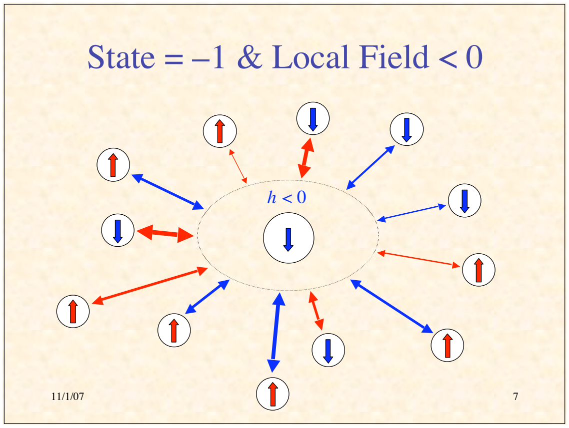
- Negative *sense* (sign)
- Large *strength*

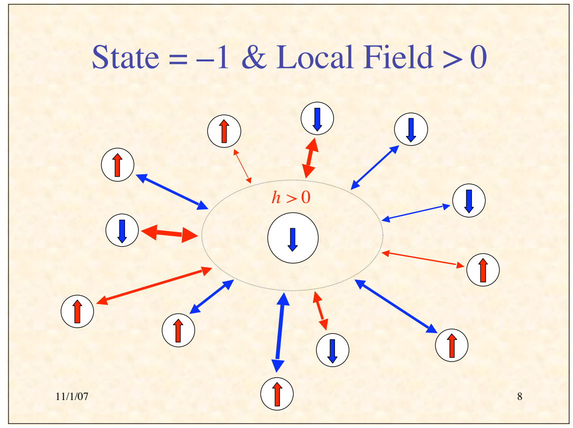


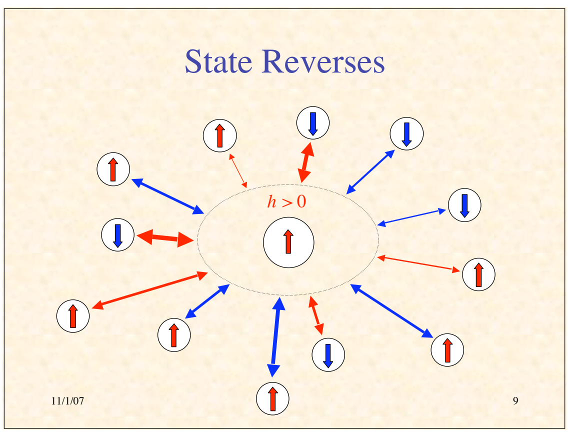
Weak Coupling

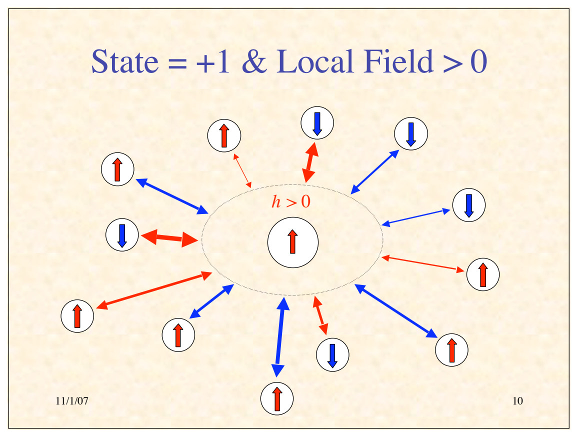
- Either *sense* (sign)
- Little *strength*

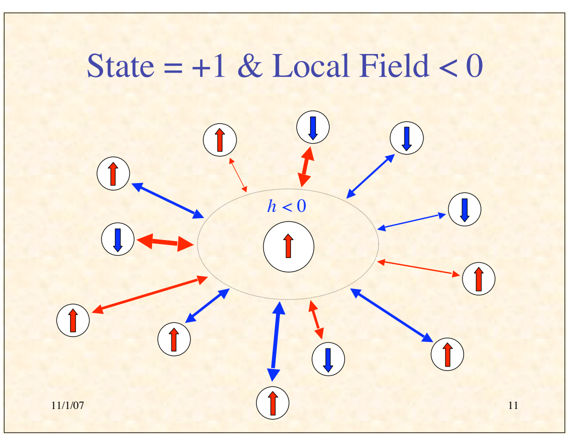


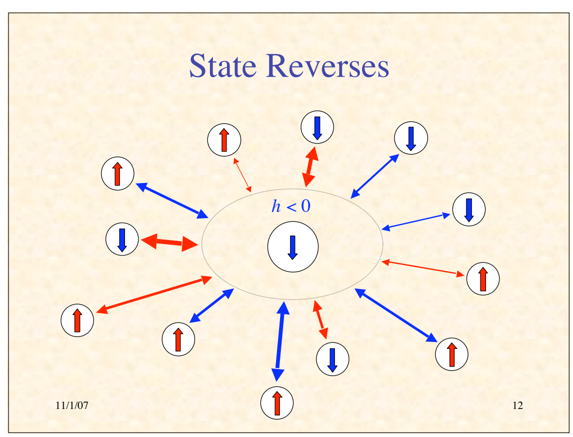

State = –1 & Local Field < 0
h < 0
11/1/07
7


State = –1 & Local Field > 0
h > 0
11/1/07
8


State Reverses
h > 0
11/1/07
9


State = +1 & Local Field > 0
h > 0
11/1/07
10


State = +1 & Local Field < 0
h < 0
11/1/07
11


State Reverses
h < 0
11/1/07
12

NetLogo Demonstration of Hopfield State Updating

Run Hopfield-update.nlogo

Hopfield Net as Soft Constraint Satisfaction System

- States of neurons as yes/no decisions
- Weights represent *soft constraints* between decisions
  - *hard* constraints *must* be respected
  - *soft* constraints have *degrees* of importance
- Decisions change to better respect constraints
- Is there an optimal set of decisions that best respects all constraints?

Demonstration of Hopfield Net Dynamics I

Run Hopfield-dynamics.nlogo

Demonstration of Hopfield Net Dynamics II

Run initialized Hopfield.nlogo

Convergence

- Does such a system converge to a stable state?
- Under what conditions does it converge?
- There is a sense in which each step relaxes the "tension" in the system
- But could a relaxation of one neuron lead to greater tension in other places?

Quantifying "Tension"

- If $w_{ij} > 0$, then s_i and s_j want to have the same sign ($s_i s_j = +1$)
- If $w_{ij} < 0$, then s_i and s_j want to have opposite signs ($s_i s_j = -1$)
- If $w_{ij} = 0$, their signs are independent
- Strength of interaction varies with $|w_{ij}|$
- Define disharmony ("tension") D_{ij} between neurons i and j:

  $D_{ij} = - s_i w_{ij} s_j$
  $D_{ij} < 0 \Rightarrow$ they are happy
  $D_{ij} > 0 \Rightarrow$ they are unhappy

Total Energy of System

The "energy" of the system is the total "tension" (disharmony) in it:

$$E\{\mathbf{s}\} = \sum_{\langle ij \rangle} D_{ij} = -\sum_{\langle ij \rangle} s_i w_{ij} s_j$$
$$= -\tfrac{1}{2}\sum_i \sum_{j \neq i} s_i w_{ij} s_j$$
$$= -\tfrac{1}{2}\sum_i \sum_j s_i w_{ij} s_j$$
$$= -\tfrac{1}{2}\mathbf{s}^{\mathrm{T}}\mathbf{W}\mathbf{s}$$

11/1/07 19

Review of Some Vector Notation

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}^{\mathrm{T}}$$ (column vectors)

$$\mathbf{x}^{\mathrm{T}}\mathbf{y} = \sum_{i=1}^{n} x_i y_i = \mathbf{x} \cdot \mathbf{y}$$ (inner product)

$$\mathbf{x}\mathbf{y}^{\mathrm{T}} = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & \ddots & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{pmatrix}$$ (outer product)

$$\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{y} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i M_{ij} y_j$$ (quadratic form)

11/1/07 20

Another View of Energy

The energy measures the number of neurons whose states are in disharmony with their local fields (i.e. of opposite sign):

$$E\{\mathbf{s}\} = -\tfrac{1}{2}\sum_i \sum_j s_i w_{ij} s_j$$
$$= -\tfrac{1}{2}\sum_i s_i \sum_j w_{ij} s_j$$
$$= -\tfrac{1}{2}\sum_i s_i h_i$$
$$= -\tfrac{1}{2}\mathbf{s}^{\mathrm{T}}\mathbf{h}$$

11/1/07 21

Do State Changes Decrease Energy?

- Suppose that neuron k changes state
- Change of energy:

$$\Delta E = E\{\mathbf{s}'\} - E\{\mathbf{s}\}$$
$$= -\sum_{\langle ij \rangle} s'_i w_{ij} s'_j + \sum_{\langle ij \rangle} s_i w_{ij} s_j$$
$$= -\sum_{j \neq k} s'_k w_{kj} s_j + \sum_{j \neq k} s_k w_{kj} s_j$$
$$= -(s'_k - s_k)\sum_{j \neq k} w_{kj} s_j$$
$$= -\Delta s_k h_k$$
$$< 0$$

11/1/07 22

Energy Does Not Increase

- In each step in which a neuron is considered for update:
  $E\{\mathbf{s}(t+1)\} - E\{\mathbf{s}(t)\} \le 0$
- Energy cannot increase
- Energy decreases if any neuron changes
- Must it stop?

11/1/07 23

Proof of Convergence in Finite Time

- There is a minimum possible energy:
  - The number of possible states $\mathbf{s} \in \{-1, +1\}^n$ is finite
  - Hence $E_{\min} = \min \{E(\mathbf{s}) \mid \mathbf{s} \in \{\pm 1\}^n\}$ exists
- Must show it is reached in a finite number of steps

11/1/07 24

Steps are of a Certain Minimum Size

If $h_k > 0$, then (let $h_{\min}$ = min of possible positive h)

$$h_k \geq \min\left\{ h \,\middle|\, h = \sum_{j \neq k} w_{kj} s_j \wedge \mathbf{s} \in \{\pm 1\}^n \wedge h > 0 \right\} =_{\text{df}} h_{\min}$$

$$\Delta E = -\Delta s_k h_k = -2h_k \leq -2h_{\min}$$

If $h_k < 0$, then (let $h_{\max}$ = max of possible negative h)

$$h_k \geq \max\left\{ h \,\middle|\, h = \sum_{j \neq k} w_{kj} s_j \wedge \mathbf{s} \in \{\pm 1\}^n \wedge h < 0 \right\} =_{\text{df}} h_{\max}$$

$$\Delta E = -\Delta s_k h_k = 2h_k \leq 2h_{\max}$$

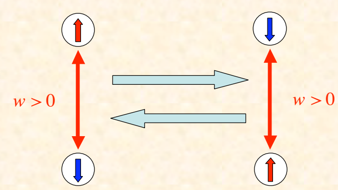
Conclusion

- If we do asynchronous updating, the Hopfield net must reach a stable, minimum energy state in a finite number of updates
- This does not imply that it is a global minimum

Lyapunov Functions

- A way of showing the convergence of discrete- or continuous-time dynamical systems
- For discrete-time system:
  - need a Lyapunov function E ("energy" of the state)
  - E is bounded below ($E\{\mathbf{s}\} > E_{\min}$)
  - $\Delta E < (\Delta E)_{\max} \leq 0$ (energy decreases a certain minimum amount each step)
  - then the system will converge in finite time
- Problem: finding a suitable Lyapunov function

Example Limit Cycle with Synchronous Updating

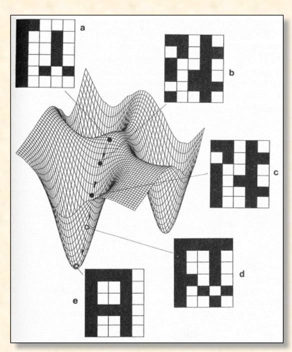


The Hopfield Energy Function is Even

- A function f is **odd** if $f(-x) = -f(x)$, for all x
- A function f is **even** if $f(-x) = f(x)$, for all x
- Observe:

$$E\{-\mathbf{s}\} = -\tfrac{1}{2}(-\mathbf{s})^{\mathrm{T}}\mathbf{W}(-\mathbf{s}) = -\tfrac{1}{2}\mathbf{s}^{\mathrm{T}}\mathbf{W}\mathbf{s} = E\{\mathbf{s}\}$$

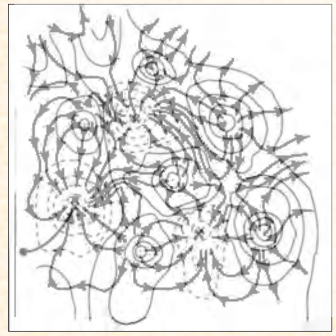
Conceptual Picture of Descent on Energy Surface



(fig. from Solé & Goodwin)

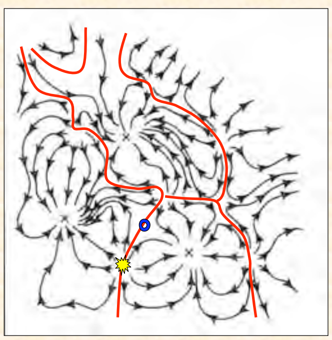
Energy Surface

11/1/07 (fig. from Haykin *Neur. Netw.*) 31



Energy Surface + Flow Lines

11/1/07 (fig. from Haykin *Neur. Netw.*) 32



Flow Lines

Basins of Attraction

11/1/07 (fig. from Haykin *Neur. Netw.*) 33