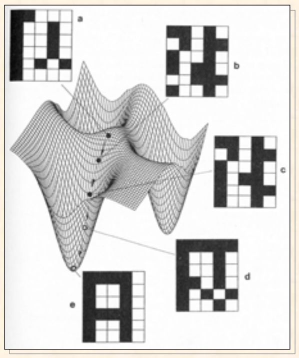
Storing Memories as Attractors



(fig. from Solé & Goodwin)

Demonstration of Hopfield Net

Run Hopfield Demo

Example of Pattern Restoration

(fig. from Arbib 1995)

Example of Pattern Restoration

(fig. from Arbib 1995)

Example of Pattern Restoration

(fig. from Arbib 1995)

11/8/04 5

Example of Pattern Restoration

(fig. from Arbib 1995)

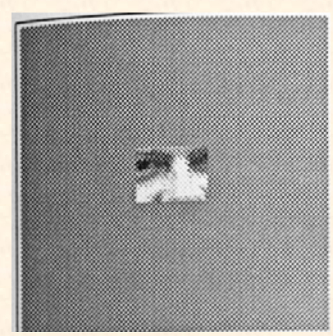
11/8/04 6

Example of Pattern Restoration

(fig. from Arbib 1995)

11/8/04 7

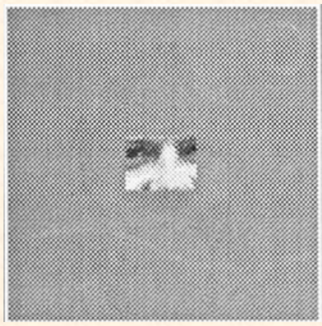
Example of Pattern Completion



(fig. from Arbib 1995)

11/8/04 8

Example of Pattern Completion



(fig. from Arbib 1995)

11/8/04 9

Example of Pattern Completion

(fig. from Arbib 1995)

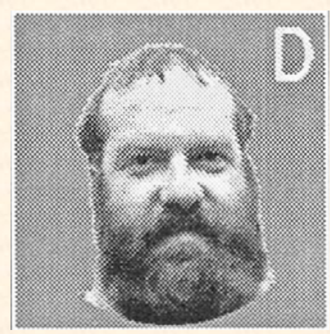
11/8/04 10

Example of Pattern Completion

(fig. from Arbib 1995)

11/8/04 11

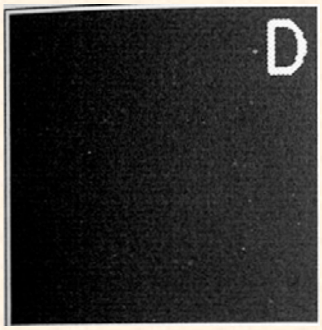
Example of Pattern Completion



(fig. from Arbib 1995)

11/8/04 12

Example of Association



(fig. from Arbib 1995)

Example of Association

(fig. from Arbib 1995)

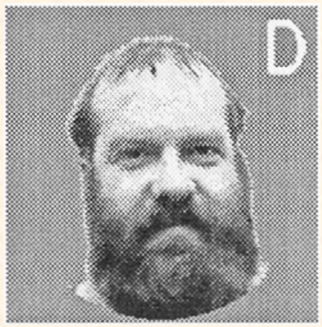
Example of Association

(fig. from Arbib 1995)

Example of Association

(fig. from Arbib 1995)

Example of Association



(fig. from Arbib 1995)

Applications of Hopfield Memory

- Pattern restoration
- Pattern completion
- Pattern generalization
- Pattern association

Hopfield Net for Optimization and for Associative Memory

- For optimization:
  - we know the weights (couplings)
  - we want to know the minima (solutions)
- For associative memory:
  - we know the minima (retrieval states)
  - we want to know the weights

Demonstration of Hopfield Net

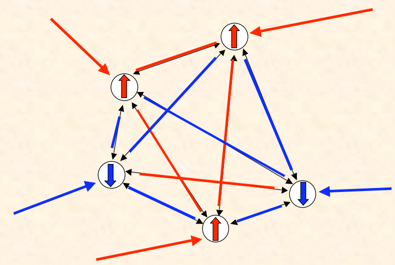
Run Hopfield Demo

Hebb's Rule

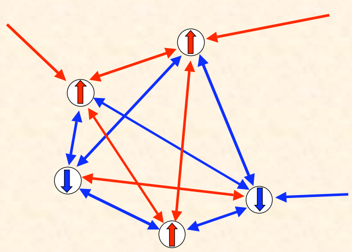
"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

—Donald Hebb (*The Organization of Behavior*, 1949, p. 62)

Example of Hebbian Learning: Pattern Imprinted



Example of Hebbian Learning: Partial Pattern Reconstruction



Mathematical Model of Hebbian Learning for One Pattern

$$\text{Let } W_{ij} = \begin{cases} x_i x_j, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}$$

$$\text{Since } x_i x_i = x_i^2 = 1, \quad \mathbf{W} = \mathbf{x}\mathbf{x}^{\mathrm{T}} - \mathbf{I}$$

For simplicity, we will include self-coupling:

$$\mathbf{W} = \mathbf{x}\mathbf{x}^{\mathrm{T}}$$

A Single Imprinted Pattern is a Stable State

- Suppose $\mathbf{W} = \mathbf{x}\mathbf{x}^{\mathrm{T}}$
- Then $\mathbf{h} = \mathbf{W}\mathbf{x} = \mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{x} = n\mathbf{x}$ since
$$\mathbf{x}^{\mathrm{T}}\mathbf{x} = \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (\pm 1)^2 = n$$
- Hence, if initial state is $\mathbf{s} = \mathbf{x}$, then new state is $\mathbf{s}' = \mathrm{sgn}\,(n\,\mathbf{x}) = \mathbf{x}$
- May be other stable states (e.g., $-\mathbf{x}$)

Questions

- How big is the basin of attraction of the imprinted pattern?
- How many patterns can be imprinted?
- Are there unneeded *spurious* stable states?
- These issues will be addressed in the context of multiple imprinted patterns

Imprinting Multiple Patterns

- Let $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p$ be patterns to be imprinted
- Define the sum-of-outer-products matrix:

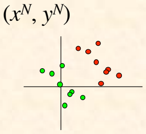
$$W_{ij} = \tfrac{1}{n} \sum_{k=1}^{p} x_i^k x_j^k$$

$$\mathbf{W} = \tfrac{1}{n} \sum_{k=1}^{p} \mathbf{x}^k \left(\mathbf{x}^k\right)^{\mathrm{T}}$$

Definition of Covariance

Consider samples $(x^1, y^1), (x^2, y^2), \ldots, (x^N, y^N)$

Let $\bar{x} = \langle x^k \rangle$ and $\bar{y} = \langle y^k \rangle$



Covariance of x and y values:

$$\begin{aligned}
C_{xy} &= \left\langle \left(x^k - \bar{x}\right)\left(y^k - \bar{y}\right) \right\rangle \\
&= \left\langle x^k y^k - \bar{x} y^k - x^k \bar{y} + \bar{x} \cdot \bar{y} \right\rangle \\
&= \left\langle x^k y^k \right\rangle - \bar{x}\left\langle y^k \right\rangle - \left\langle x^k \right\rangle \bar{y} + \bar{x} \cdot \bar{y} \\
&= \left\langle x^k y^k \right\rangle - \bar{x} \cdot \bar{y} - \bar{x} \cdot \bar{y} + \bar{x} \cdot \bar{y} \\
C_{xy} &= \left\langle x^k y^k \right\rangle - \bar{x} \cdot \bar{y}
\end{aligned}$$

Weights & the Covariance Matrix

Sample pattern vectors: $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^p$

Covariance of i^{th} and j^{th} components:

$$C_{ij} = \langle x_i^k x_j^k \rangle - \overline{x_i} \cdot \overline{x_j}$$

If $\forall i : \overline{x_i} = 0$ (±1 equally likely in all positions):

$$C_{ij} = \langle x_i^k x_j^k \rangle = \tfrac{1}{p} \sum_{k=1}^{p} x_i^k y_j^k$$

$$\therefore \mathbf{W} = \tfrac{p}{n} \mathbf{C}$$

Characteristics of Hopfield Memory

- Distributed (“holographic”)
  – every pattern is stored in every location (weight)
- Robust
  – correct retrieval in spite of noise or error in patterns
  – correct operation in spite of considerable weight damage or noise

Stability of Imprinted Memories

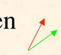
- Suppose the state is one of the imprinted patterns $\mathbf{x}^m$
- Then: $\mathbf{h} = \mathbf{W}\mathbf{x}^m = \left[\tfrac{1}{n}\sum_k \mathbf{x}^k (\mathbf{x}^k)^{\mathrm{T}}\right]\mathbf{x}^m$

$$= \tfrac{1}{n}\sum_k \mathbf{x}^k (\mathbf{x}^k)^{\mathrm{T}} \mathbf{x}^m$$

$$= \tfrac{1}{n}\mathbf{x}^m (\mathbf{x}^m)^{\mathrm{T}} \mathbf{x}^m + \tfrac{1}{n}\sum_{k \neq m} \mathbf{x}^k (\mathbf{x}^k)^{\mathrm{T}} \mathbf{x}^m$$

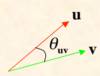
$$= \mathbf{x}^m + \tfrac{1}{n}\sum_{k \neq m} (\mathbf{x}^k \cdot \mathbf{x}^m)\mathbf{x}^k$$

Interpretation of Inner Products

- $\mathbf{x}^k \cdot \mathbf{x}^m = n$ if they are identical 
  – highly correlated
- $\mathbf{x}^k \cdot \mathbf{x}^m = -n$ if they are complementary 
  – highly correlated (reversed)
- $\mathbf{x}^k \cdot \mathbf{x}^m = 0$ if they are orthogonal 
  – largely uncorrelated
- $\mathbf{x}^k \cdot \mathbf{x}^m$ measures the *crosstalk* between patterns k and m 

Cosines and Inner products

$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta_{\mathbf{uv}}$



If $\mathbf{u}$ is bipolar, then $\|\mathbf{u}\|^2 = \mathbf{u} \cdot \mathbf{u} = n$

Hence, $\mathbf{u} \cdot \mathbf{v} = \sqrt{n} \sqrt{n} \cos \theta_{\mathbf{uv}} = n \cos \theta_{\mathbf{uv}}$