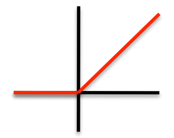
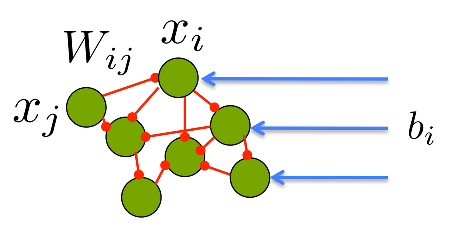
Perturbative Memory Encoding in Recurrent Networks

Carina Curto[1], Anda Degeratu[2], Vladimir Itskov[1]
[1]Department of Mathematics, University of Nebraska-Lincoln; [2]Max Planck Institute, Potsdam, Germany.

The problem: How can we encode memories as fixed point attractors of a network without incurring unwanted spurious states?

The framework: threshold-linear networks





dynamics on the network

$$\frac{dx_i}{dt} = -\frac{1}{\tau_i}x_i + \left[\sum_{j=1}^{n} W_{ij}x_j + b_i\right]_+$$

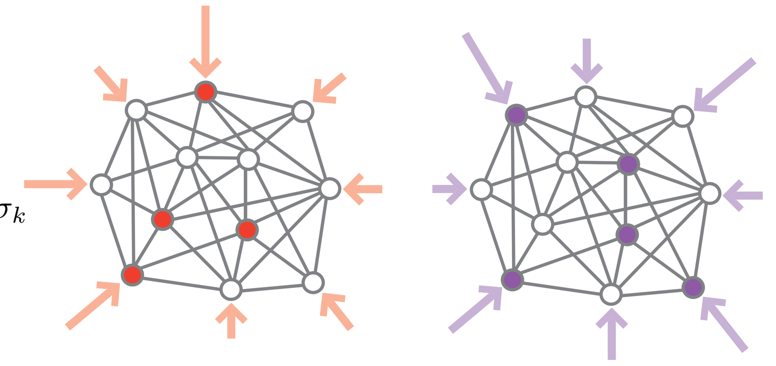
$$\dot{x} = -Dx + [Wx+b]_+$$

The inputs are allowed to vary, so the network is fully-characterized by the pair of matrices: (W, D)

- Input: a vector $b \in \mathbb{R}^n$
- Memory is recalled: The network evolves to a stable fixed point with support matching the memory pattern.

fixed point x^* $\quad \mathrm{supp}(x^*) = \{i \in \{1,\dots,n\} \mid x_i^* > 0\} = \sigma_k$

Hahnloser, Seung, and Slotine, *Neural Computation* (2003)



Binary memory patterns
$\xi^1,\dots,\xi^p \in \{0,1\}^N$
or, equivalently,
$\sigma_1,\dots,\sigma_p \subset \{1,\dots,n\}$

Memory patterns as stable sets

$$\mathcal{M}(W,D) \stackrel{\text{def}}{=} \{\sigma \subset \{1,\dots,n\} \mid \sigma \text{ is a stable set of } (W,D)\}.$$

A non-empty subset of neurons $\sigma \subset \{1,...,n\}$ is a *stable set* of (W,D) if there exists an asymptotically stable fixed point x^* such that $\mathrm{supp}(x^*) = \sigma$, for at least one external input vector $b \in \mathbb{R}^n$. An *unstable set* of (W,D) is a non-empty subset of neurons that is not stable.

Problem: Given a prescribed list of memory patterns, $\sigma_1,\dots,\sigma_p \subset \{1,\dots,n\}$ can we find a synaptic matrix such that these - and only these – memories are encoded?

Key observation:

$$\sigma \text{ is a stable set} \iff (-D+W)_\sigma \text{ is a stable principal submatrix}$$

A principal submatrix is obtained by restricting to a subset of indices:

Curto, Degeratu, Itskov, *Bulletin of Mathematical Biology* (2011)

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \qquad A_{\{124\}} = \begin{pmatrix} a_{11} & a_{12} & a_{14} \\ a_{21} & a_{22} & a_{24} \\ a_{41} & a_{42} & a_{44} \end{pmatrix}$$

A matrix is stable if all its eigenvalues have negative real part.

Translation of the problem:

1. For a given set of memory patterns

$$\sigma_1,\dots,\sigma_p \subset \{1,\dots,n\}$$

can we find a matrix

$$M = -D + W$$

such that:

- M_{σ_i} is stable for $i = 1,\dots,p$
- All other principal submatrices of M are not stable.

2. For what sets of memory patterns is this possible?

A perturbative approach:

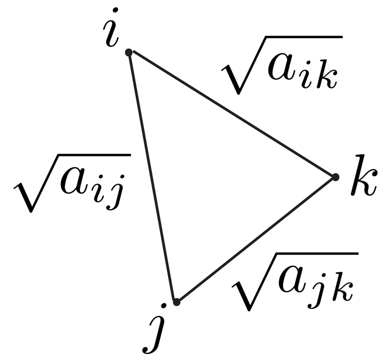
$$M = -11^T + \varepsilon A$$

where $\varepsilon > 0$ and A is a Hebbian matrix.

A Hebbian matrix A is a real-valued $n \times n$ matrix that is symmetric, has non-negative entries, and zeroes along the diagonal. $\qquad A_{ij} = A_{ji} \geq 0$ and $A_{ii} = 0$

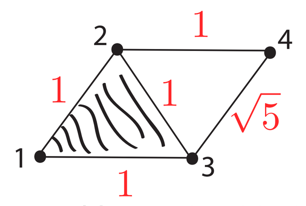
A geometric observation:

$$M_{\{ijk\}} = \begin{pmatrix} -1 & -1+a_{ij} & -1+a_{ik} \\ -1+a_{ij} & -1 & -1+a_{jk} \\ -1+a_{ik} & -1+a_{jk} & -1 \end{pmatrix}$$



$$M_{\{ijk\}} \text{ is stable} \iff \sqrt{a_{ij}}, \sqrt{a_{jk}} \text{ and } \sqrt{a_{ik}} \text{ can be edge lengths of a non degenerate triangle}$$

An example on 4 neurons



max stable sets: {123},{24},{34}
min unstable sets: {234},{14}

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 5 \\ 0 & 1 & 5 & 0 \end{pmatrix}$$

$$M = -11^T + \varepsilon A = \begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix} + \varepsilon \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 5 \\ 0 & 1 & 5 & 0 \end{pmatrix}$$

$M_{\{123\}}, M_{\{24\}}$, and $M_{\{34\}}$ are all stable for $0 < \varepsilon < 0.4$

$M_{\{234\}}, M_{\{14\}}$ unstable for all $\varepsilon > 0$

This generalizes! Main geometric result:

Theorem. Let $\varepsilon > 0$ and A an $n \times n$ Hebbian matrix for $n \geq 2$. Then the perturbed matrix $-11^T + \varepsilon A$ is a stable matrix if and only if the following two conditions hold:

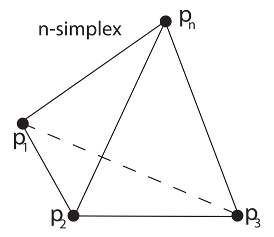
1. A is a nondegenerate square distance matrix, and
2. $0 < \varepsilon < -\dfrac{\mathcal{C}(A)}{\det A}$, where $\mathcal{C}(A)$ is the Cayley-Menger determinant of A.

$$\mathcal{C}(A) \stackrel{\text{def}}{=} \det \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & A \end{bmatrix}$$

A non-degenerate square distance matrix A is an $n \times n$ matrix with entries

$$A_{ij} = \|p_i - p_j\|^2,$$

where $p_1, .., p_n$ are the vertices of a nondegenerate n-simplex in $\mathbb{R}^{n-1}$.



The geometric result allows us to construct synaptic matrices for many prescribed lists of memory patterns.

Theorem. Let G be a graph on n vertices, and $p_1, ..., p_n \in \mathbb{R}^d$ a configuration of points in Euclidean space. Consider the matrix $M = -11^T + \varepsilon A$ for $\varepsilon > 0$ and

$$A_{ij} = \begin{cases} \|p_i - p_j\|^2, & \text{for } (ij) \in G, \\ \leq 0, & \text{for } (ij) \notin G. \end{cases}$$

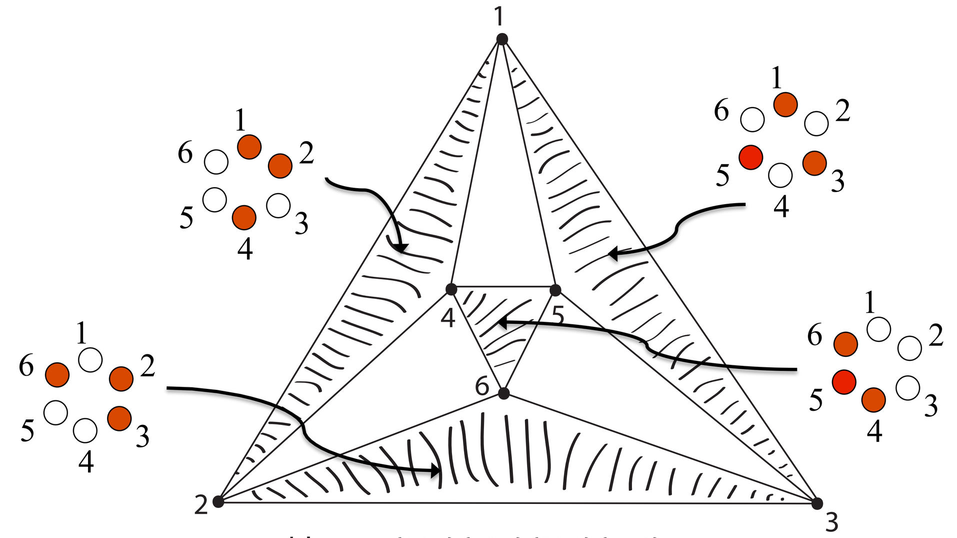
Then, for any $\sigma \subseteq [n]$ such that $|\sigma| \geq 2$,

$$M_\sigma \text{ is stable} \iff \begin{cases} 1. & \{p_i\}_{i\in\sigma} \text{ are in general position, and} \\ 2. & |\sigma| \leq d+1 \text{ and } \sigma \text{ is a clique of } G, \text{ and} \\ 3. & 0 < \varepsilon < -\mathcal{C}(A_\sigma)/\det A_\sigma. \end{cases}$$

Note that if $|\sigma| = 1$, then $M_\sigma = (-1)$ is stable.

An example on 6 neurons

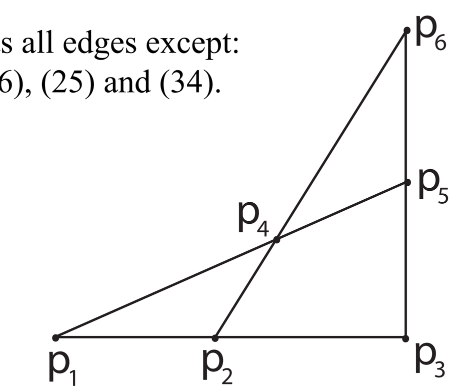
Problem:
Find a 6x6 symmetric matrix whose stable principal submatrices are given by this simplicial complex.



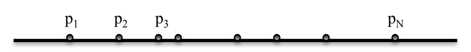
max stable sets: {124},{135},{236},{456}
min unstable sets: {123},{145},{246},{356},{16},{25},{34}

Solution:
Use the following graph and arrangement of points in the plane.

G has all edges except: (16), (25) and (34).



Encoding pairs and triples of co-active neurons



$$W_{ij} = -k + \frac{1}{N}\|p_i - p_j\|^2 + \varepsilon A_{ij}$$

By perturbing around a "line model" architecture, we can selectively encode prescribed sets of pairs and triples of co-active neurons.

Conclusions

1. Within this framework of memory encoding, we can encode sets of highly overlapping memory patterns exactly using a perturbative approach.
2. Precise encoding of overlapping memory patterns can be achieved using a geometric characterization for the stability of principal submatrices of the synaptic matrix.