8 Absence of Multistability in Recurrent Linear Networks

We consider the possibility of a network that uses a Hebb-like rule to construct a symmetric weight matrix, but for which the neurons act as linear devices, i.e., the output of the cell is a linear function of the input. This clearly is not the case for cells that vary from quiescent to spiking as a function of their input, but could be the case for cells whose spike rate is uniformly and monotonically modulated up and down. It is also the case for networks of cells with solely graded synaptic release.

Our goal is to show that such networks fail an associative memories, i.e., recurrent linear networks map all inputs onto the a single output state. For the present purpose, we focus on this issue as motivation for networks of threshold elements.

The input to the cell is given by

$$u_i = \sum_{j=1}^{N} W_{ij} S_j \tag{8.101}$$

where W_{ij} is the (symmetric) connection matrix. For a linear network, the input and output are related by

$$S_i = g u_i \tag{8.102}$$

where g is the gain. In a parallel updating scheme, in which we explicitly note the time steps, we have

$$S_i(t+1) = g \sum_{j=1}^{N} W_{ij} S_j(t). \tag{8.103}$$

In vector notation, this is

$$\mathbf{S}(t+1) = g\mathbf{W} \cdot \mathbf{S}(t) \tag{8.104}$$

Now we can itterate, the synchronous equivalent of recurrence, starting from time $t = 0$:

$$\begin{aligned}
\mathbf{S}(1) &= g\mathbf{W} \cdot \mathbf{S}(0) \\
\mathbf{S}(2) &= g\mathbf{W} \cdot \mathbf{S}(1) \\
\mathbf{S}(3) &= g\mathbf{W} \cdot \mathbf{S}(2) \\
&\;\;. \\
&\;\;. \\
\mathbf{S}(n+1) &= g\mathbf{W} \cdot \mathbf{S}(n)
\end{aligned} \tag{8.105}$$

This becomes

$$\mathbf{S}(n) = g^n \mathbf{W}^n \cdot \mathbf{S}(0) \tag{8.106}$$

Now we recall that $\mathbf{W}$ satisfies an eigenvalue equation;

$$\mathbf{W} \cdot \mathbf{V}_k = \lambda_k \mathbf{V}_k \tag{8.107}$$

where k labels labels the eigenvalue and $1 < k < N$. We can rotate the symmetric matrix matrix $\mathbf{W}$ by a unity transformation that preserves the eignevalues, i.e,

$$\mathbf{W} = \mathbf{U \Lambda U^T} \tag{8.108}$$

where $\mathbf{U}$ is a unitary matrix defined through $\mathbf{U} \cdot \mathbf{U^T} = \mathbf{1}$. The diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues along the diagonal, such that

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & \\ 0 & 0 & \lambda_3 & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \end{pmatrix}$$

Clearly the rotated eigenvectors, $\mathbf{U^T V}$, are of the form

$$\mathbf{U^T V}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \qquad \mathbf{U^T V}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \cdots$$

since $\mathbf{W} \cdot \mathbf{V}_k = \lambda_k \mathbf{V}_k$ implies $\mathbf{\Lambda} \cdot \mathbf{U^T V}_k = \lambda_k \mathbf{U^T V}_k$ so the $\mathbf{U^T V}_k$ are the eigenvalues of the diagonalized (rotated) system.

Now we can go back to the iterative expression for $\mathbf{S}(n)$.

$$\begin{aligned} \mathbf{S}(n) &= g^n \mathbf{W}^n \cdot \mathbf{S}(0) \\ &= g^n \left(\mathbf{U \Lambda U^T}\right)^n \cdot \mathbf{S}(0) \\ &= g^n \mathbf{U \Lambda^n U^T} \cdot \mathbf{S}(0) \end{aligned} \tag{8.109}$$

where we used

$$\begin{aligned} \left(\mathbf{U \Lambda U^T}\right)^n &= \mathbf{U \Lambda U^T U \Lambda U^T} \cdots \mathbf{U \Lambda U^T} \\ &= \mathbf{U \Lambda^n U^T} \end{aligned} \tag{8.110}$$

Thus

$$\mathbf{U^T S}(n) = g^n \mathbf{\Lambda^n} \cdot \mathbf{U^T S}(0) \tag{8.111}$$

But the diagonal matrix $\mathbf{\Lambda^n}$, when rank ordered so that λ_1 is the dominant eigenvalue, becomes,

$$\mathbf{\Lambda}^n = \begin{pmatrix} \lambda_1^n & 0 & 0 & \cdots \\ 0 & \lambda_2^n & 0 & \\ 0 & 0 & \lambda_3^n & \\ . & & & \\ . & & & \\ . & & & \end{pmatrix} = \lambda_1^n \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & \left(\frac{\lambda_2}{\lambda_1}\right)^n & 0 & \\ 0 & 0 & \left(\frac{\lambda_3}{\lambda_1}\right)^n & \\ . & & & \\ . & & & \\ . & & & \end{pmatrix} \to \lambda_1^n \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \\ . & & & \\ . & & & \\ . & & & \end{pmatrix}$$

Thus the system converges to the dominant eigenvector, $\mathbf{S}_1 = \mathbf{UV}_1$, and eigenvalue, λ_1, independent of the initial starting state. Thus only a single state is supported in an iterative network with linear neurons. In other words

$$\mathbf{S}(n) \to (g\lambda_1)^n \, \mathbf{S}_1 \tag{8.112}$$

As a practical issue, the gain must be chosen to keep the output within some practical limits or the amplitude must be renomalized afer teach update. Nonetheless, the essential issue is that neurons that function as linear transducers are NOT usable as associative networks.

We will next consider the lessons from analyzing networks of neurons that function as threshold devices. McCulloch and Pitts realized the potential of such networks already by 1943, and Hopfield understood the importance of threshold nonlinearities for multistability in recurrent networks in 1982. By the mid 1980's the full statistical mechanics of these networks was worked out by Sompolinsky, Amit and Gutfriend, culminating in Amit's 1989 book "Modeling Brain Function: The World of Attractor Neural Networks".