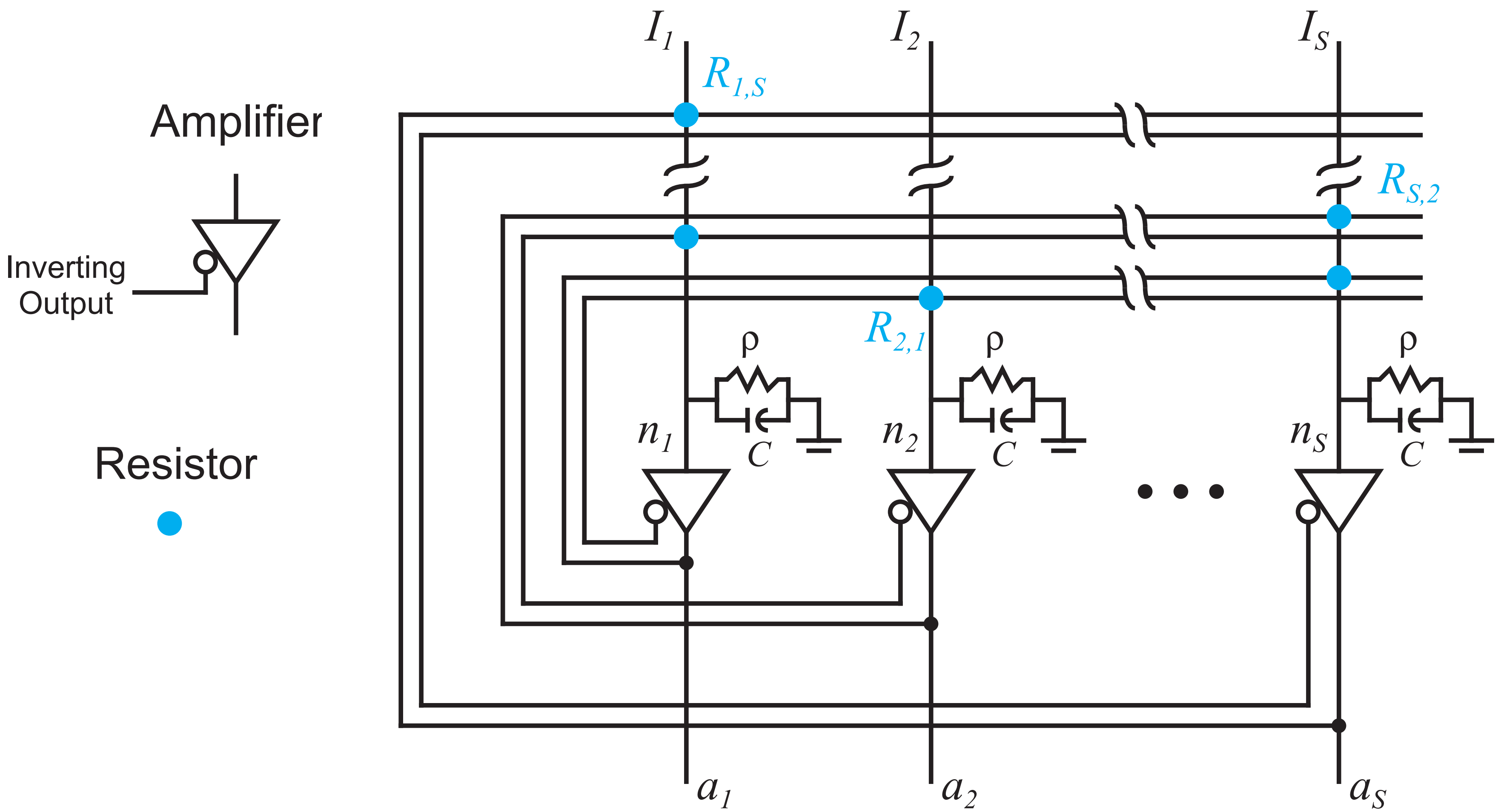
Hopfield Network

Hopfield Model



Equations of Operation

$$C\frac{dn_i(t)}{dt} = \sum_{j=1}^{S} T_{i,j}a_j(t) - \frac{n_i(t)}{R_i} + I_i$$

n_i - input voltage to the *i*th amplifier
a_i - output voltage of the *i*th amplifier
C - amplifier input capacitance
I_i - fixed input current to the *i*th amplifier

$$|T_{i,j}| = \frac{1}{R_{i,j}} \qquad \frac{1}{R_i} = \frac{1}{\rho} + \sum_{j=1}^{S}\frac{1}{R_{i,j}} \qquad n_i = f^{-1}(a_i) \qquad a_i = f(n_i)$$

Network Format

$$R_i C \frac{dn_i(t)}{dt} = \sum_{j=1}^{S} R_i T_{i,j} a_j(t) - n_i(t) + R_i I_i$$

Define:

$$\varepsilon = R_i C \qquad w_{i,j} = R_i T_{i,j} \qquad b_i = R_i I_i$$

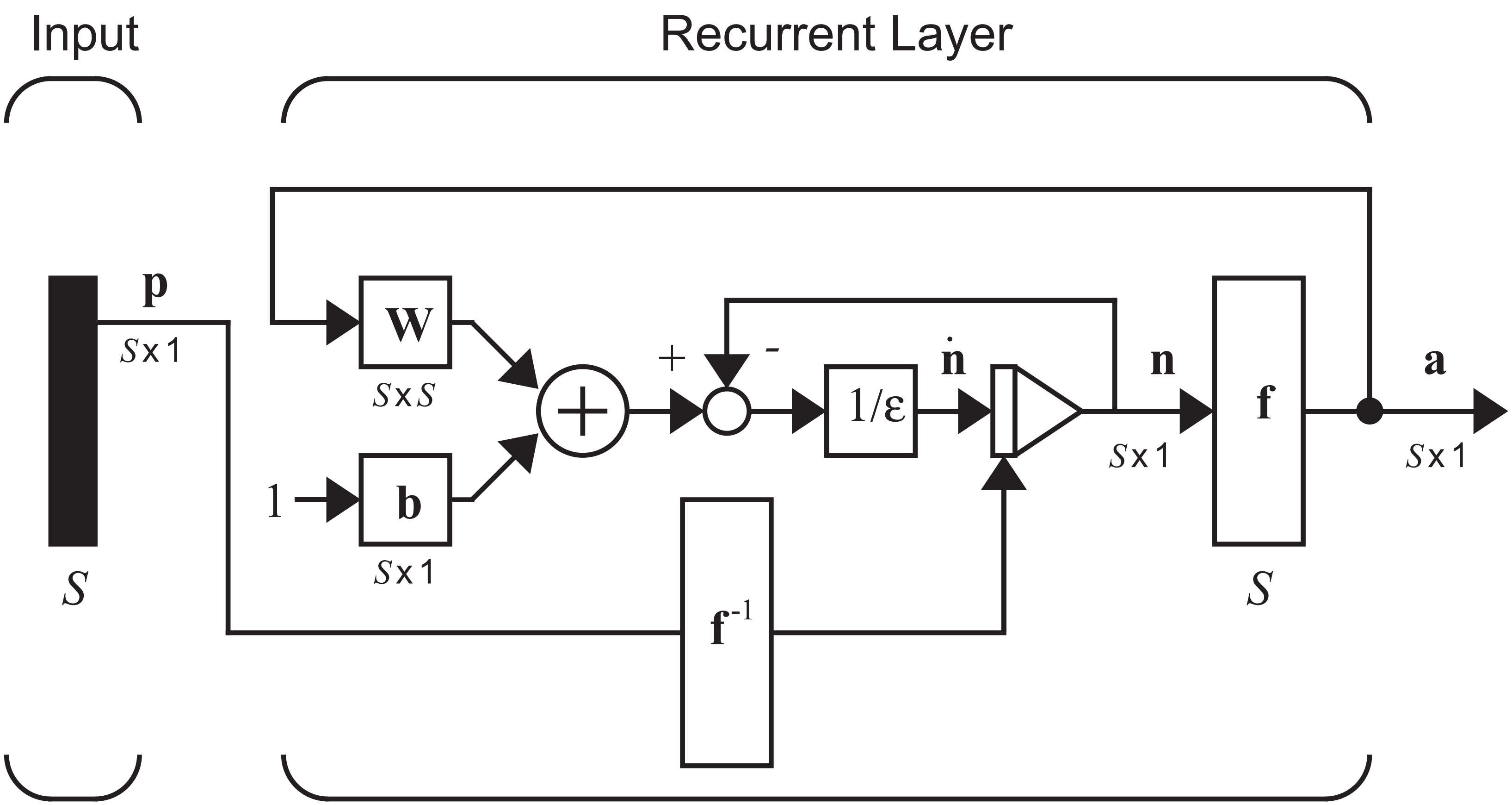
$$\varepsilon \frac{dn_i(t)}{dt} = -n_i(t) + \sum_{j=1}^{S} w_{i,j} a_j(t) + b_i$$

Vector Form:

$$\varepsilon \frac{d\mathbf{n}(t)}{dt} = -\mathbf{n}(t) + \mathbf{W}\mathbf{a}(t) + \mathbf{b}$$

$$\mathbf{a}(t) = \mathbf{f}(\mathbf{n}(t))$$

Hopfield Network



$$\mathbf{n}(0) = \mathbf{f}^{-1}(\mathbf{p}), \quad (\mathbf{a}(0) = \mathbf{p}) \qquad \varepsilon\, d\mathbf{n}/dt = -\mathbf{n} + \mathbf{W}\,\mathbf{f}(\mathbf{n}) + \mathbf{b}$$

Lyapunov Function

$$V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} + \sum_{i=1}^{S}\left\{\int_0^{a_i} f^{-1}(u)\,du\right\} - \mathbf{b}^T\mathbf{a}$$

Individual Derivatives

First Term:

$$\frac{d}{dt}\left\{-\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a}\right\} = -\frac{1}{2}\nabla[\mathbf{a}^T\mathbf{W}\mathbf{a}]^T\frac{d\mathbf{a}}{dt} = -[\mathbf{W}\mathbf{a}]^T\frac{d\mathbf{a}}{dt} = -\mathbf{a}^T\mathbf{W}\frac{d\mathbf{a}}{dt}$$

Second Term:

$$\frac{d}{dt}\left\{\int_0^{a_i} f^{-1}(u)\,du\right\} = \frac{d}{da_i}\left\{\int_0^{a_i} f^{-1}(u)\,du\right\}\frac{da_i}{dt} = f^{-1}(a_i)\frac{da_i}{dt} = n_i\frac{da_i}{dt}$$

$$\frac{d}{dt}\left[\sum_{i=1}^{S}\left\{\int_0^{a_i} f^{-1}(u)\,du\right\}\right] = \mathbf{n}^T\frac{d\mathbf{a}}{dt}$$

Third Term:

$$\frac{d}{dt}\{-\mathbf{b}^T\mathbf{a}\} = -\nabla[\mathbf{b}^T\mathbf{a}]^T\frac{d\mathbf{a}}{dt} = -\mathbf{b}^T\frac{d\mathbf{a}}{dt}$$

21

Complete Lyapunov Derivative

$$\frac{d}{dt}V(\mathbf{a}) = -\mathbf{a}^T\mathbf{W}\frac{d\mathbf{a}}{dt} + \mathbf{n}^T\frac{d\mathbf{a}}{dt} - \mathbf{b}^T\frac{d\mathbf{a}}{dt} = [-\mathbf{a}^T\mathbf{W} + \mathbf{n}^T - \mathbf{b}^T]\frac{d\mathbf{a}}{dt}$$

From the system equations we know:

$$[-\mathbf{a}^T\mathbf{W} + \mathbf{n}^T - \mathbf{b}^T] = -\varepsilon\left[\frac{d\mathbf{n}(t)}{dt}\right]^T$$

So the derivative can be written:

$$\frac{d}{dt}V(\mathbf{a}) = -\varepsilon\left[\frac{d\mathbf{n}(t)}{dt}\right]^T\frac{d\mathbf{a}}{dt} = -\varepsilon\sum_{i=1}^{S}\left(\frac{dn_i}{dt}\right)\left(\frac{da_i}{dt}\right) = -\varepsilon\sum_{i=1}^{S}\left(\frac{dn_i}{dt}\right)\left(\frac{da_i}{dt}\right)$$

$$= -\varepsilon\sum_{i=1}^{S}\left(\frac{d}{da_i}[f^{-1}(a_i)]\right)\left(\frac{da_i}{dt}\right)^2$$

If $\frac{d}{da_i}[f^{-1}(a_i)] > 0$ then $\frac{d}{dt}V(\mathbf{a}) \leq 0$

Invariant Sets

$$Z = \{\mathbf{a}: dV(\mathbf{a})/dt = 0,\ \mathbf{a} \text{ in the closure of } G\}$$

$$\frac{d}{dt}V(\mathbf{a}) = -\varepsilon \sum_{i=1}^{S} \left(\frac{d}{da_i}[f^{-1}(a_i)]\right)\left(\frac{da_i}{dt}\right)^2$$

This will be zero only if the neuron outputs are not changing:

$$\frac{d\mathbf{a}}{dt} = \mathbf{0}$$

Therefore, the system energy is not changing only at the equilibrium points of the circuit. Thus, all points in Z are potential attractors:

$$L = Z$$

Example

$$a = f(n) = \frac{2}{\pi}\tan^{-1}\left(\frac{\gamma\pi n}{2}\right) \qquad n = \frac{2}{\gamma\pi}\tan\left(\frac{\pi}{2}a\right)$$

$$\left.\begin{array}{l} R_{1,2} = R_{2,1} = 1 \\ \\ T_{1,2} = T_{2,1} = 1 \end{array}\right\} \quad \mathbf{W} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\varepsilon = R_i C = 1$$

$$\gamma = 1.4$$

$$\left. I_1 = I_2 = 0 \right\} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

21

Example Lyapunov Function

$$V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} + \sum_{i=1}^{S}\left\{\int_0^{a_i} f^{-1}(u)\,du\right\} - \mathbf{b}^T\mathbf{a}$$

$$-\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} = -\frac{1}{2}\begin{bmatrix} a_1 & a_2 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = -a_1 a_2$$

$$\int_0^{a_i} f^{-1}(u)\,du = \frac{2}{\gamma\pi}\int_0^{a_i} \tan\left(\frac{\pi}{2}u\right)du = \frac{2}{\gamma\pi}\left[-\log\left[\cos\left(\frac{\pi}{2}u\right)\right]\frac{2}{\pi}\right]_0^{a_i} = -\frac{4}{\gamma\pi^2}\log\left[\cos\left(\frac{\pi}{2}a_i\right)\right]$$

$$V(\mathbf{a}) = -a_1 a_2 - \frac{4}{1.4\pi^2}\left[\log\left\{\cos\left(\frac{\pi}{2}a_1\right)\right\} + \log\left\{\cos\left(\frac{\pi}{2}a_2\right)\right\}\right]$$

Example Network Equations

$$\frac{d\mathbf{n}}{dt} = -\mathbf{n} + \mathbf{W}\mathbf{f}(\mathbf{n}) = -\mathbf{n} + \mathbf{W}\mathbf{a}$$

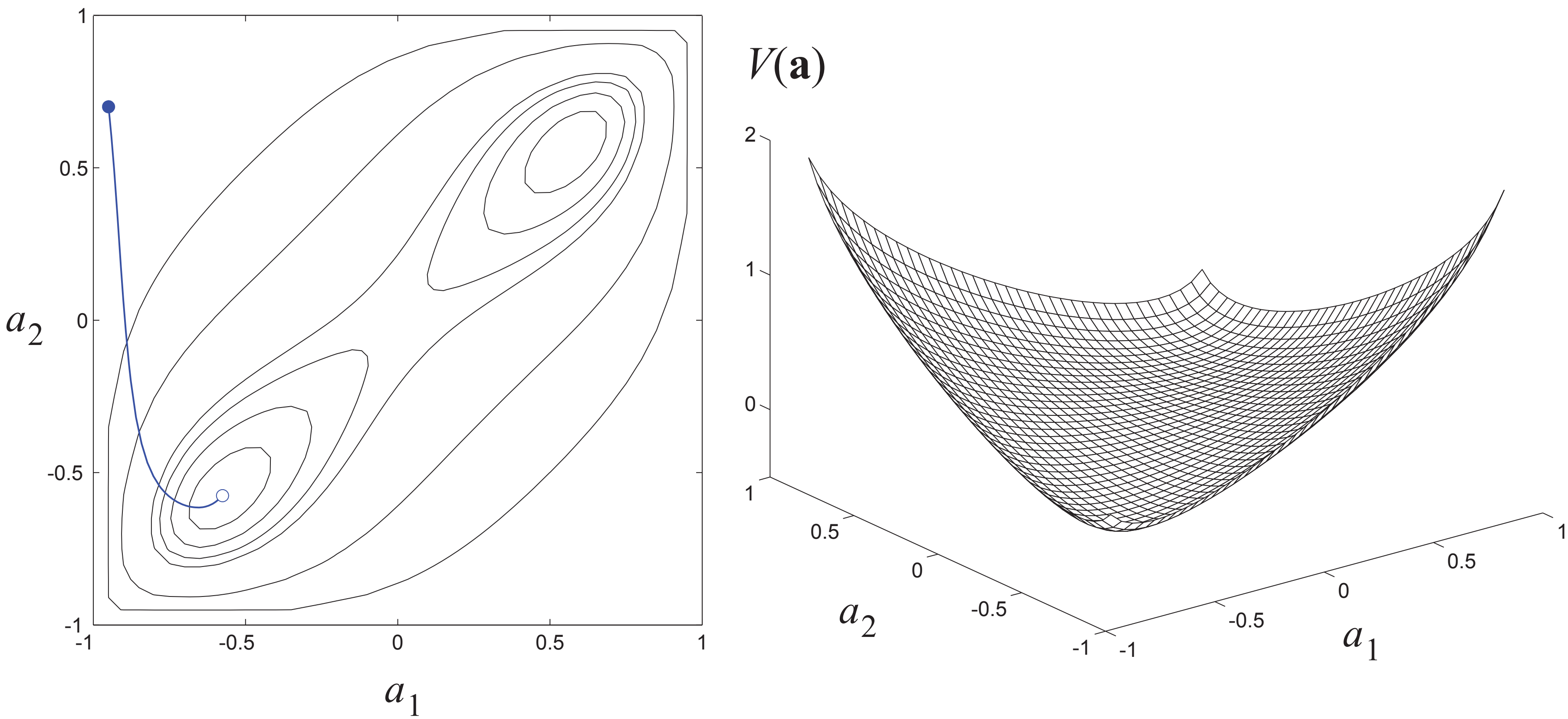
$$dn_1/dt = a_2 - n_1$$

$$dn_2/dt = a_1 - n_2$$

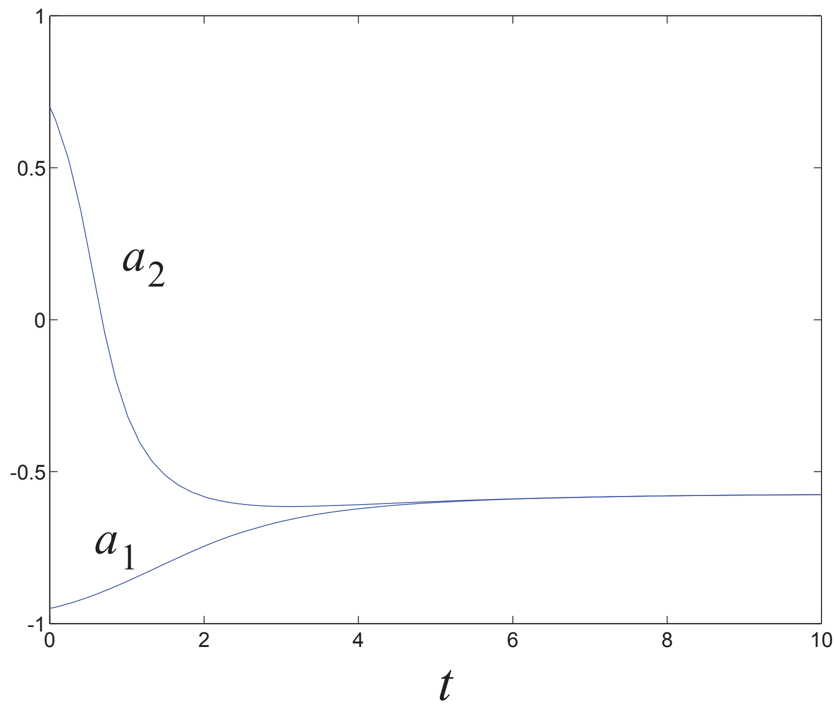
$$a_1 = \frac{2}{\pi}\tan^{-1}\left(\frac{1.4\pi}{2}n_1\right)$$

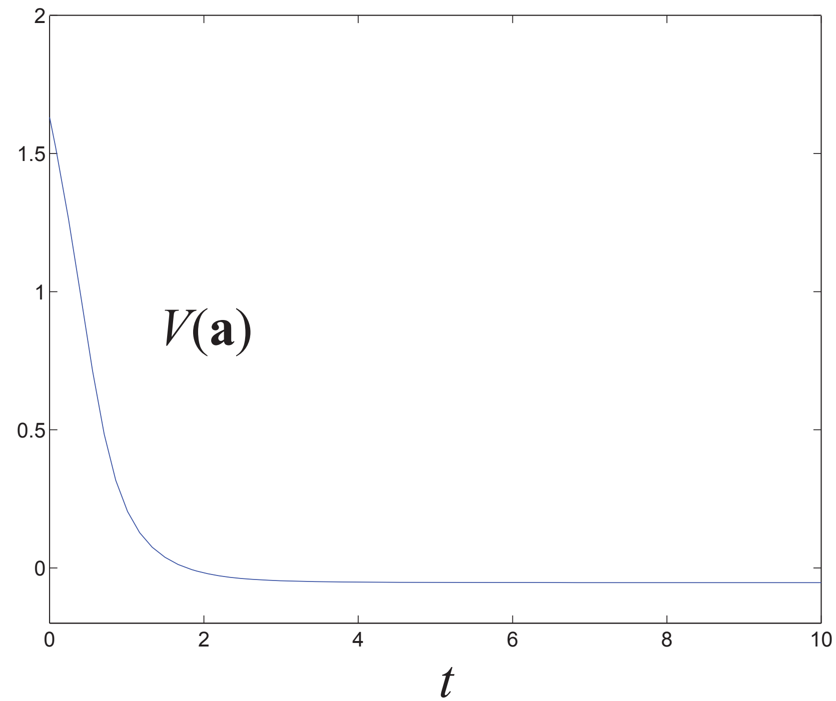
$$a_2 = \frac{2}{\pi}\tan^{-1}\left(\frac{1.4\pi}{2}n_2\right)$$

21 Lyapunov Function and Trajectory

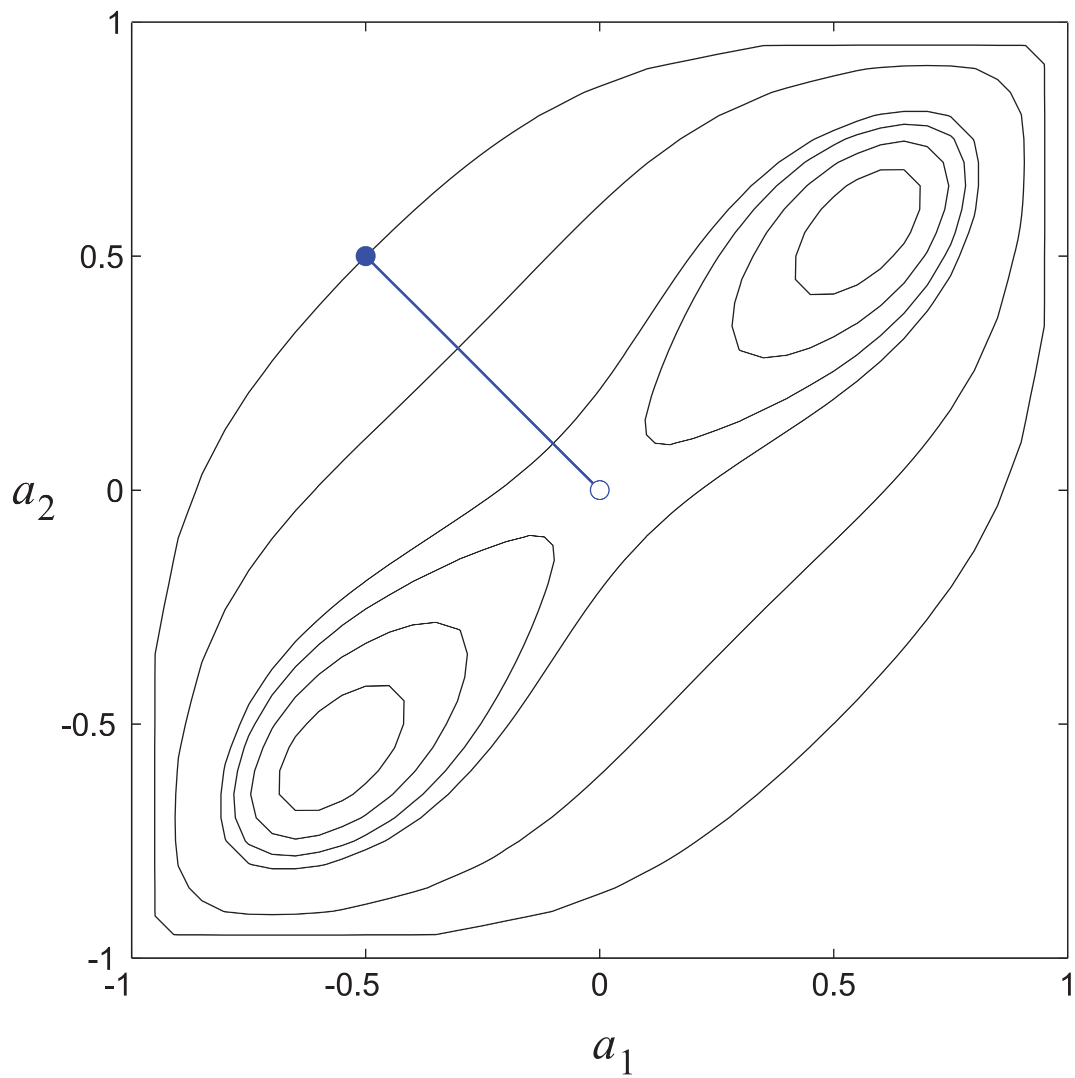


Time Response





Convergence to a Saddle Point



Hopfield Attractors

The potential attractors of the Hopfield network satisfy:

$$\frac{d\mathbf{a}}{dt} = \mathbf{0}$$

How are these points related to the minima of $V(\mathbf{a})$? The minima must satisfy:

$$\nabla V = \left[\frac{\partial V}{\partial a_1}\ \frac{\partial V}{\partial a_2}\ \cdots\ \frac{\partial V}{\partial a_S}\right]^T = \mathbf{0}$$

Where the Lyapunov function is given by:

$$V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} + \sum_{i=1}^{S}\left\{\int_0^{a_i} f^{-1}(u)du\right\} - \mathbf{b}^T\mathbf{a}$$

21

Hopfield Attractors

Using previous results, we can show that:

$$\nabla V(\mathbf{a}) = [-\mathbf{W}\mathbf{a} + \mathbf{n} - \mathbf{b}] = -\varepsilon\left[\frac{d\mathbf{n}(t)}{dt}\right]$$

The *i*th element of the gradient is therefore:

$$\frac{\partial}{\partial a_i} V(\mathbf{a}) = -\varepsilon \frac{dn_i}{dt} = -\varepsilon \frac{d}{dt}([f^{-1}(a_i)]) = -\varepsilon \frac{d}{d\,a_i}[f^{-1}(a_i)]\frac{da_i}{d\,t}$$

Since the transfer function and its inverse are monotonic increasing:

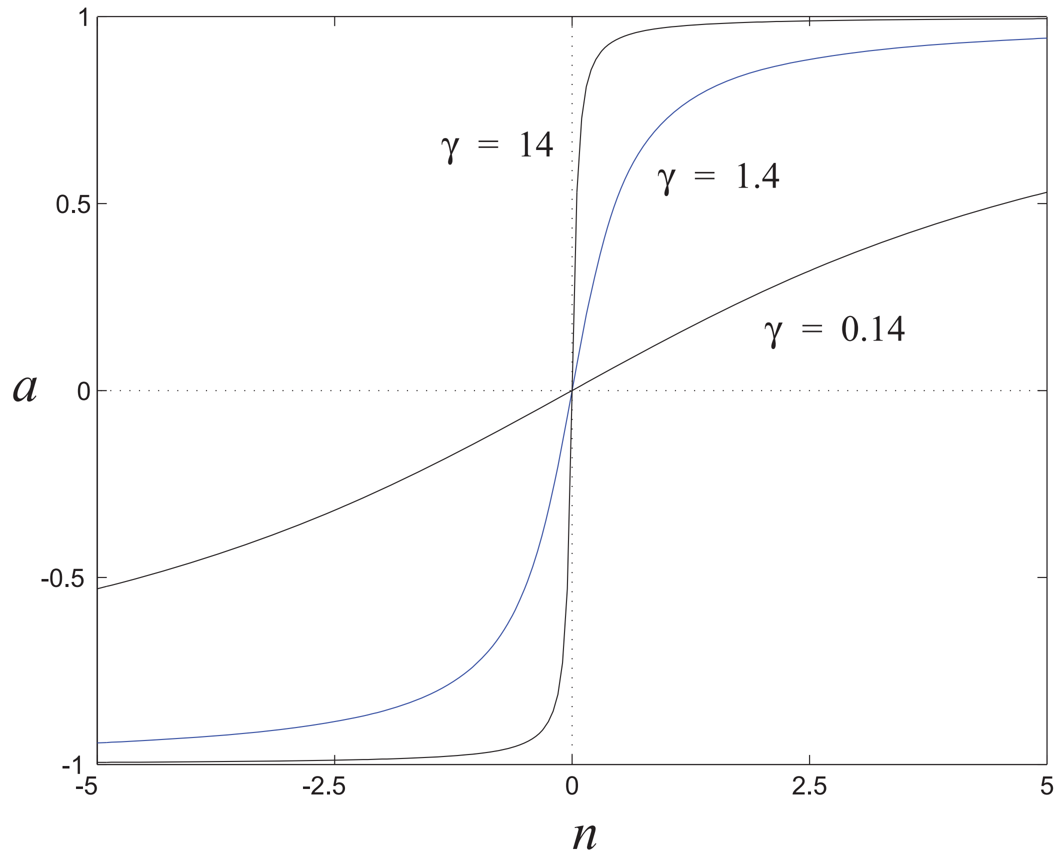
$$\frac{d}{da_i}[f^{-1}(a_i)] > 0$$

All points for which $\frac{d\mathbf{a}(t)}{dt} = \mathbf{0}$ will also satisfy $\nabla V(\mathbf{a}) = \mathbf{0}$

Therefore all attractors will be stationary points of $V(\mathbf{a})$.

Effect of Gain

$$a = f(n) = \frac{2}{\pi}\tan^{-1}\left(\frac{\gamma\pi n}{2}\right)$$



21

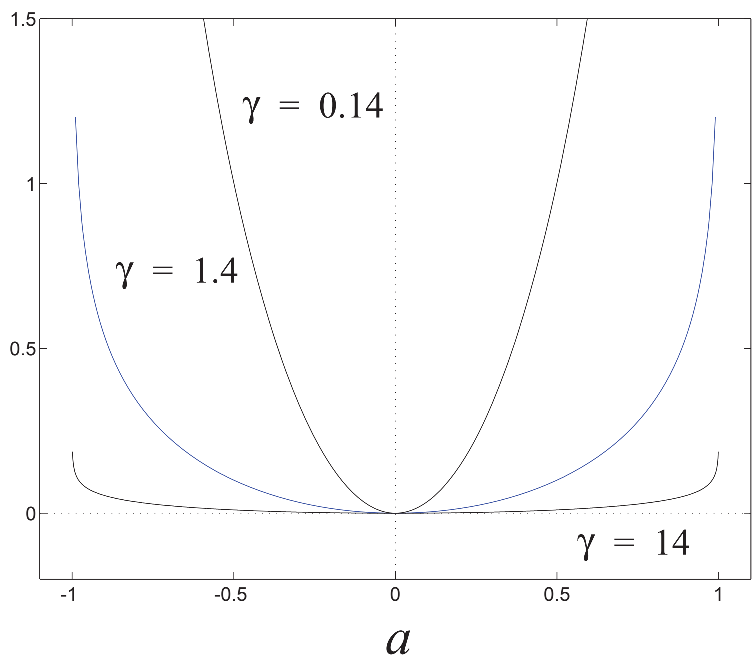
Lyapunov Function

$$V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} + \sum_{i=1}^{S}\left\{\int_0^{a_i} f^{-1}(u)\,du\right\} - \mathbf{b}^T\mathbf{a}$$

$$f^{-1}(u) = \frac{2}{\gamma\pi}\tan\left(\frac{\pi u}{2}\right)$$

$$\int_0^{a_i} f^{-1}(u)\,du = \frac{2}{\gamma\pi}\left[\frac{2}{\pi}\log\left(\cos\left(\frac{\pi a_i}{2}\right)\right)\right] = -\frac{4}{\gamma\pi^2}\log\left[\cos\left(\frac{\pi a_i}{2}\right)\right]$$

$$-\frac{4}{\gamma\pi^2}\log\left[\cos\left(\frac{\pi a}{2}\right)\right]$$



21 High Gain Lyapunov Function

As $\gamma \rightarrow \infty$ the Lyapunov function reduces to:

$$V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} - \mathbf{b}^T\mathbf{a}$$

The high gain Lyapunov function is quadratic:

$$V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} - \mathbf{b}^T\mathbf{a} = \frac{1}{2}\mathbf{a}^T\mathbf{A}\mathbf{a} + \mathbf{d}^T\mathbf{a} + c$$

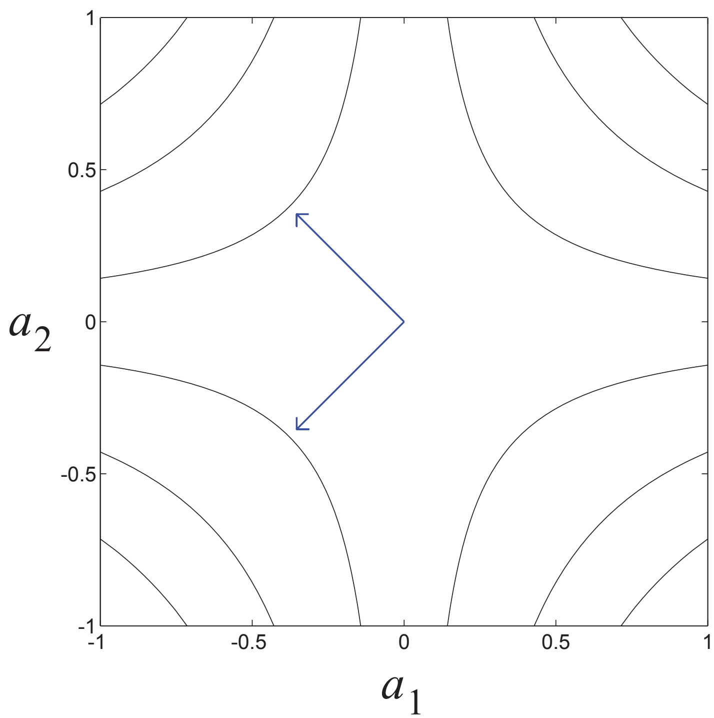
where

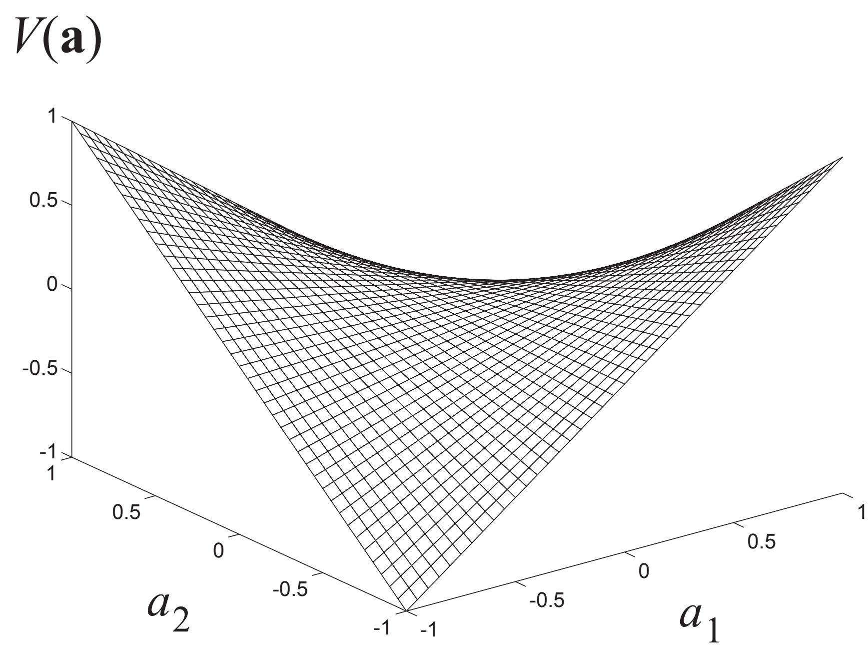
$$\nabla^2 V(\mathbf{a}) = \mathbf{A} = -\mathbf{W} \qquad \mathbf{d} = -\mathbf{b} \qquad c = 0$$

Example

$$\nabla^2 V(\mathbf{a}) = -\mathbf{W} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \qquad \left|\nabla^2 V(\mathbf{a}) - \lambda \mathbf{I}\right| = \begin{vmatrix} -\lambda & -1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 - 1 = (\lambda + 1)(\lambda - 1)$$

$$\lambda_1 = -1 \qquad \mathbf{z}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad\qquad \lambda_2 = 1 \qquad \mathbf{z}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$





Hopfield Design

The Hopfield network will minimize the following Lyapunov function:

$$V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} - \mathbf{b}^T\mathbf{a}$$

Choose the weight matrix **W** and the bias vector **b** so that V takes on the form of a function you want to minimize.

21 Content-Addressable Memory

Content-Addressable Memory - retrieves stored memories on the basis of part of the contents.

Prototype Patterns:

$$\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_Q\} \qquad \text{(bipolar vectors)}$$

Proposed Performance Index:

$$J(\mathbf{a}) = -\frac{1}{2}\sum_{q=1}^{Q}\left([\mathbf{p}_q]^T\mathbf{a}\right)^2$$

For orthogonal prototypes, if we evaluate the performance index at a prototype:

$$J(\mathbf{p}_j) = -\frac{1}{2}\sum_{q=1}^{Q}\left([\mathbf{p}_q]^T\mathbf{p}_j\right)^2 = -\frac{1}{2}\left([\mathbf{p}_j]^T\mathbf{p}_j\right)^2 = -\frac{S}{2}$$

$J(\mathbf{a})$ will be largest when **a** is not close to any prototype pattern, and smallest when **a** is equal to a prototype pattern.

Hebb Rule

If we use the supervised Hebb rule to compute the weight matrix:

$$\mathbf{W} = \sum_{q=1}^{Q} \mathbf{p}_q (\mathbf{p}_q)^T \qquad \mathbf{b} = \mathbf{0}$$

the Lyapunov function will be:

$$V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} = -\frac{1}{2}\mathbf{a}^T\left[\sum_{q=1}^{Q} \mathbf{p}_q (\mathbf{p}_q)^T\right]\mathbf{a} = -\frac{1}{2}\sum_{q=1}^{Q} \mathbf{a}^T\mathbf{p}_q (\mathbf{p}_q)^T\mathbf{a}$$

This can be rewritten:

$$V(\mathbf{a}) = -\frac{1}{2}\sum_{q=1}^{Q} [(\mathbf{p}_q)^T\mathbf{a}]^2 = J(\mathbf{a})$$

Therefore the Lyapunov function is equal to our performance index for the content addressable memory.

Hebb Rule Analysis

$$\mathbf{W} = \sum_{q=1}^{Q} \mathbf{p}_q (\mathbf{p}_q)^T$$

If we apply prototype $\mathbf{p}_j$ to the network:

$$\mathbf{W}\mathbf{p}_j = \sum_{q=1}^{Q} \mathbf{p}_q (\mathbf{p}_q)^T \mathbf{p}_j = \mathbf{p}_j (\mathbf{p}_j)^T \mathbf{p}_j = S\mathbf{p}_j$$

Therefore each prototype is an eigenvector, and they have a common eigenvalue of S. The eigenspace for the eigenvalue $\lambda = S$ is therefore:

$$X = \mathrm{span}\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_Q\}$$

An S-dimensional space of all vectors which can be written as linear combinations of the prototype vectors.

21

Weight Matrix Eigenspace

The entire input space can be divided into two disjoint sets:

$$R^S = X \cup X^{\perp}$$

where $X^{\perp}$ is the orthogonal complement of X. For vectors **a** in the orthogonal complement we have:

$$(\mathbf{p}_q)^T \mathbf{a} = 0, \quad q = 1, 2, \ldots, Q$$

Therefore,

$$\mathbf{W}\mathbf{a} = \sum_{q=1}^{Q} \mathbf{p}_q (\mathbf{p}_q)^T \mathbf{a} = \sum_{q=1}^{Q} (\mathbf{p}_q \times 0) = \mathbf{0} = 0 \times \mathbf{a}$$

The eigenvalues of **W** are S and 0, with corresponding eigenspaces of X and $X^{\perp}$. For the Hessian matrix

$$\nabla^2 V = -\mathbf{W}$$

the eigenvalues are $-S$ and 0, with the same eigenspaces.

Lyapunov Surface

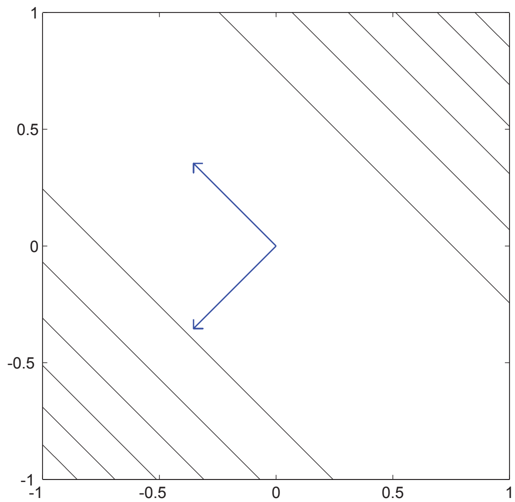
The high-gain Lyapunov function is a quadratic function. Therefore, the eigenvalues of the Hessian matrix determine its shape. Because the first eigenvalue is negative, V will have negative curvature in X. Because the second eigenvalue is zero, V will have zero curvature in $X^{\perp}$.

Because V has negative curvature in X, the trajectories of the Hopfield network will tend to fall into the corners of the hypercube $\{\mathbf{a}: -1 < a_i < 1\}$ that are contained in X.

Example

21

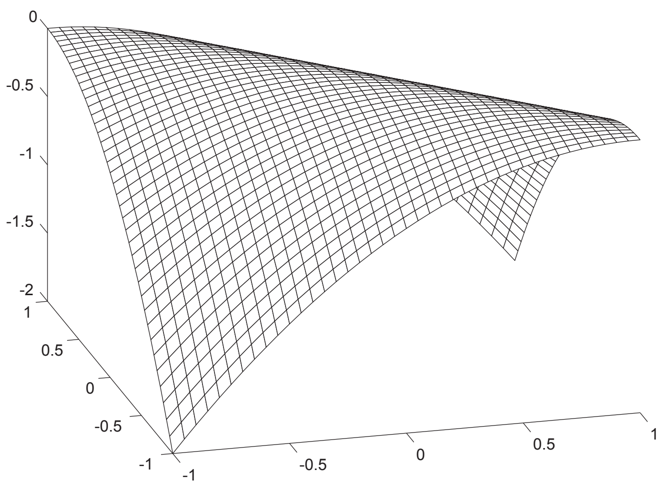
$$\mathbf{p}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{W} = \mathbf{p}_1(\mathbf{p}_1)^T = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad V(\mathbf{a}) = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} = -\frac{1}{2}\mathbf{a}^T\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\mathbf{a}$$



$$\nabla^2 V(\mathbf{a}) = -\mathbf{W} = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}$$

$$\lambda_1 = -S = -2 \qquad \mathbf{z}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$X = \{\mathbf{a} : a_1 = a_2\}$$



$$\lambda_2 = 0 \qquad \mathbf{z}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$X^{\perp} = \{\mathbf{a} : a_1 = -a_2\}$$

21 Zero Diagonal Elements

We can zero the diagonal elements of the weight matrix:

$$\mathbf{W}' = \mathbf{W} - Q\mathbf{I}$$

The prototypes remain eigenvectors of this new matrix, but the corresponding eigenvalue is now (S-Q):

$$\mathbf{W}'\mathbf{p}_q = [\mathbf{W} - Q\mathbf{I}]\mathbf{p}_q = S\mathbf{p}_q - Q\mathbf{p}_q = (S - Q)\mathbf{p}_q$$

The elements of $X^{\perp}$ also remain eigenvectors of this new matrix, with a corresponding eigenvalue of (-Q):

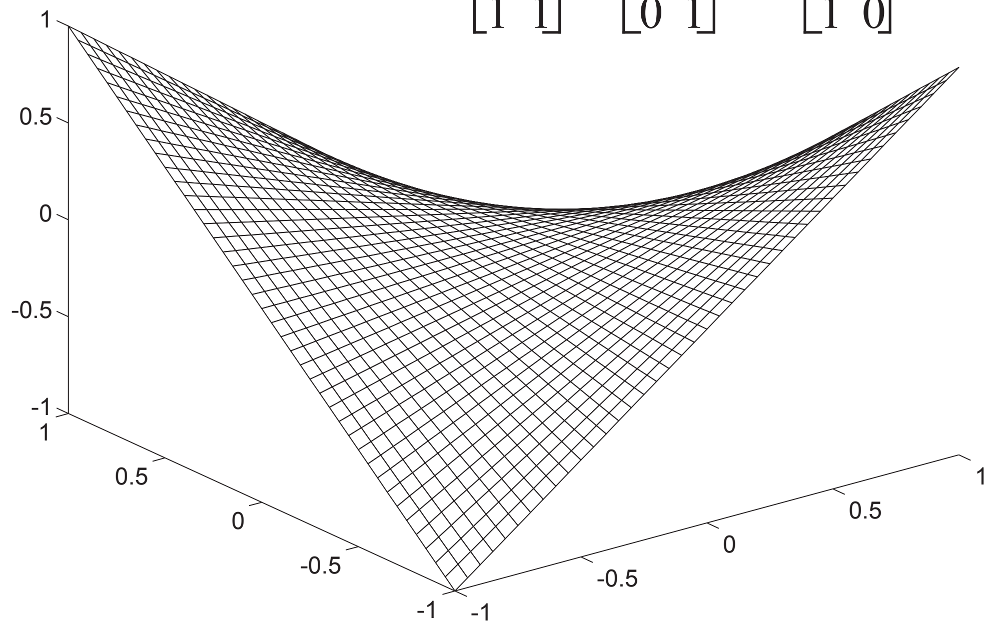
$$\mathbf{W}'\mathbf{a} = [\mathbf{W} - Q\mathbf{I}]\mathbf{a} = \mathbf{0} - Q\mathbf{a} = -Q\mathbf{a}$$

The Lyapunov surface will have negative curvature in X and positive curvature in $X^{\perp}$, in contrast with the original Lyapunov function, which had negative curvature in X and zero curvature in $X^{\perp}$.

21

Example

$$\mathbf{W}' = \mathbf{W} - Q\mathbf{I} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$



If the initial condition falls exactly on the line a_1=-a_2, and the weight matrix **W** is used, then the network output will remain constant. If the initial condition falls exactly on the line a_1=-a_2, and the weight matrix **W**' is used, then the network output will converge to the saddle point at the origin.