

Notes on Objective Function paper(Intrator and Cooper, 1992)

2.1 Experimental Results

Summary of Experimental Results	
Experiment	Reference
Monocular Deprivation	<ul style="list-style-type: none">• OD changes were observed as early as 6 h(Freeman and Olson, 1982; Mioche and Singer, 1989)• complete loss of response to closed as early as 12 h(Mioche and Singer, 1989)• moderate increase of response to the normal eye occasionally(Mioche and Singer, 1989)
Binocular Deprivation	<ul style="list-style-type: none">• cortical response reduced within 3 d(Freeman et al., 1981)
Reverse Suture	<ul style="list-style-type: none">• the time course for the reduction of response to the newly deprived eye was similar to monocular deprivation(Mioche and Singer, 1989)• At least 24 h of reverse suture is required before the responses to the deprived eye reappears(Mioche and Singer, 1989)

3 Feature Extraction

- **curse of dimensionality**: high dimensional spaces are **sparse** which leads to the fact that absurdly large amounts of training data are necessary to get low variance estimators \Rightarrow need to do **dimensionality reduction** or **feature extraction**.
- **supervised methods** for feature extraction use **global constraints** which are sensitive to number of parameters, and get caught in local minima.
- **unsupervised methods** use **local objective function** (function to be minimized, ie. energy function), so may not be so sensitive to number of parameters.
- **exploratory projection pursuit**: method for dimensionality reduction, seeking **interesting projections**
 - when one projects high dim \rightarrow low dim, most projections are **gaussian**
 - **interesting projections** \equiv projections which are far from gaussian (whatever that might mean)
 - deviations from gaussian often defined using **polynomial moments**
 - second order statistics (PCA) is not enough to characterize the important features
 - polynomial moments emphasize the **tail** of the distribution

2.2 BCM Theory

Notation	
input vector:	\mathbf{x}
weight vector of i th neuron:	\mathbf{m}_i
output of i th neuron:	c_i
learning rate:	μ
inhibition between i th to j th neuron:	$(\mathbf{L})_{ij} \equiv L_{ij}$

Modification:

$$\begin{aligned}\mathbf{m}_i &= \phi(c_i, \theta_i)\mathbf{x} \\ \theta_i &\equiv E[c_i^2]\end{aligned}$$

$$c_i = \underbrace{\mathbf{m}_i \cdot \mathbf{x}}_{\text{LGN cells}} + \underbrace{\sum_j L_{ij} c_j}_{\text{cortical cells}}$$

2.3 Mean Field Approximation

- The mean field approximation is that each neuron sees an **averaged activity** from its neighbors; that the spatial inhomogeneity in the network activity can be ignored at the local level

$$\begin{aligned}
 c_i &= \mathbf{m}_i \cdot \mathbf{x} + \sum_j L_{ij} c_j \\
 \left\{ \sum_j L_{ij} c_j \approx \bar{c} \sum_j L_{ij}, \bar{c} \equiv \frac{1}{N} \sum_i c_i \right\} &\Rightarrow c_i \approx \mathbf{m}_i \cdot \mathbf{x} + \bar{c} \sum_j L_{ij} \\
 \bar{c} &= \frac{1}{N} \sum_i c_i \\
 &= \frac{1}{N} \sum_i (\mathbf{m}_i \cdot \mathbf{x}) + \frac{1}{N} \sum_i \left(\bar{c} \sum_j L_{ij} \right) \\
 \left\{ L_o \equiv \frac{1}{N} \sum_{ij} L_{ij}, \bar{\mathbf{m}} \equiv \frac{1}{N} \sum_i \mathbf{m}_i \right\} &\Rightarrow \bar{c} = \bar{\mathbf{m}} \cdot \mathbf{x} + \bar{c} L_o \\
 &= (1 - L_o)^{-1} \bar{\mathbf{m}} \cdot \mathbf{x} \\
 c_i &= \mathbf{m}_i \cdot \mathbf{x} + (1 - L_o)^{-1} \bar{\mathbf{m}} \cdot \mathbf{x} \sum_j L_{ij} \\
 \left\{ \text{assume } L_{ij} = L_{ji} \Rightarrow L_o = \frac{1}{N} \sum_{ij} L_{ij} = \sum_i L_{ij} \right\} &\Rightarrow c_i = \mathbf{m}_i \cdot \mathbf{x} + (1 - L_o)^{-1} \bar{\mathbf{m}} \cdot \mathbf{x} L_o \\
 \left\{ \alpha \equiv -L_o(1 - L_o)^{-1} \bar{\mathbf{m}} \right\} &\Rightarrow c_i = (\mathbf{m}_i - \alpha) \cdot \mathbf{x}
 \end{aligned}$$

- One can justify **negative weights** by performing the transformation: $(\mathbf{m}_i - \alpha) \rightarrow \mathbf{m}_i$.
- what is α ?

$$\begin{aligned}
 L_o < 0 &\Rightarrow \alpha > 0 \\
 0 < L_o < 1 &\Rightarrow \alpha < 0 \\
 L_o > 1 &\Rightarrow \alpha > 0
 \end{aligned}$$

4 Objective Function Formulation

- The goal here is to come up with an **energy function whose gradient** is identical to the **BCM modification equation**
- Then we **find the specific form** of this function for a few cases: **single linear neuron, single nonlinear neuron, network with feedforward inhibition (linear and nonlinear)**

4.1 Single Linear Neuron

- Some definitions:

$$\theta \equiv E[(\mathbf{x} \cdot \mathbf{m})^2]$$

$$\begin{aligned}\hat{\phi}(c, \theta) &\equiv c^2 - \frac{1}{2}c\theta \\ \phi(c, \theta) &\equiv c^2 - c\theta\end{aligned}$$

- we want a **loss function** (ie. energy function) which, when minimized, **seeks polynomial moments**.
- we want a **loss function** which exhibits the fact that **more nodes** \Rightarrow **more interesting**
- **Loss function:**

$$\begin{aligned}L_{\mathbf{m}}(\mathbf{x}) &\equiv -\mu \int_0^{\mathbf{x} \cdot \mathbf{m}} \hat{\phi}(s, \theta) ds \\ &= -\mu \int_0^{\mathbf{x} \cdot \mathbf{m}} \left(s^2 - \frac{1}{2} s E[(\mathbf{x} \cdot \mathbf{m})^2] \right) ds \\ &= -\mu \frac{s^3}{3} \Big|_0^{\mathbf{x} \cdot \mathbf{m}} + \mu \frac{s^2}{4} E[(\mathbf{x} \cdot \mathbf{m})^2] \Big|_0^{\mathbf{x} \cdot \mathbf{m}}\end{aligned}$$

$$L_{\mathbf{m}}(\mathbf{x}) = -\mu \left(\frac{1}{3} (\mathbf{x} \cdot \mathbf{m})^3 - \frac{1}{4} (\mathbf{x} \cdot \mathbf{m})^2 E[(\mathbf{x} \cdot \mathbf{m})^2] \right)$$

- **Risk function** (expected loss):

$$R_{\mathbf{m}}(\mathbf{x}) \equiv E[L_{\mathbf{m}}(\mathbf{x})] = -\mu E \left[\frac{1}{3} (\mathbf{x} \cdot \mathbf{m})^3 - \frac{1}{4} (\mathbf{x} \cdot \mathbf{m})^2 E[(\mathbf{x} \cdot \mathbf{m})^2] \right]$$

$$R_{\mathbf{m}}(\mathbf{x}) = -\mu \left(\frac{1}{3} E[(\mathbf{x} \cdot \mathbf{m})^3] - \frac{1}{4} E^2[(\mathbf{x} \cdot \mathbf{m})^2] \right)$$

- change \mathbf{m} to minimize $R_{\mathbf{m}}$ (**gradient descent**):

$$\begin{aligned}\frac{d\mathbf{m}}{dt} &= -\frac{\partial}{\partial \mathbf{m}} R_{\mathbf{m}} = \mu \frac{\partial}{\partial \mathbf{m}} \left(\frac{1}{3} E[(\mathbf{x} \cdot \mathbf{m})^3] - \frac{1}{4} E^2[(\mathbf{x} \cdot \mathbf{m})^2] \right) \\ &= \mu E[(\mathbf{x} \cdot \mathbf{m})^2 \mathbf{x}] - \underbrace{E[(\mathbf{x} \cdot \mathbf{m})^2]}_{\theta} E[(\mathbf{x} \cdot \mathbf{m}) \mathbf{x}] \\ &= \mu E[(c^2 - c\theta) \mathbf{x}]\end{aligned}$$

$$\frac{d\mathbf{m}}{dt} = \mu E[\phi(c, \theta) \mathbf{x}] \text{ which is identical to the BCM modification equation}$$

4.2 Single Nonlinear Neuron

- to get rid of outlier sensitivity: $c = \sigma(\mathbf{x} \cdot \mathbf{m})$
- to deal with separated clusters:
 - $c = \sigma(\mathbf{x} \cdot \mathbf{m} + \beta)$ where β could be some spontaneous activity
 - or**
 - $c = \sigma(\mathbf{x} \cdot \mathbf{m})$ where $\mathbf{x} \equiv (x_1, \dots, x_n, 1)$ and $\mathbf{m} \equiv (m_1, \dots, m_n, \beta)$

- Some definitions:

$$\begin{aligned} c &= \sigma(\mathbf{x} \cdot \mathbf{m}) \\ \theta &\equiv E[\sigma^2(\mathbf{x} \cdot \mathbf{m})] \end{aligned}$$

- Loss function:

$$\begin{aligned} L_{\mathbf{m}}(\mathbf{x}) &\equiv -\mu \int_0^{\sigma(\mathbf{x} \cdot \mathbf{m})} \hat{\phi}(s, \theta) ds \\ &= -\mu \int_0^{\sigma(\mathbf{x} \cdot \mathbf{m})} \left(s^2 - \frac{1}{2} s E[(\mathbf{x} \cdot \mathbf{m})^2] \right) ds \\ &= -\mu \frac{s^3}{3} \Big|_0^{\sigma(\mathbf{x} \cdot \mathbf{m})} + \mu \frac{s^2}{4} E[(\mathbf{x} \cdot \mathbf{m})^2] \Big|_0^{\sigma(\mathbf{x} \cdot \mathbf{m})} \end{aligned}$$

$$L_{\mathbf{m}}(\mathbf{x}) = -\mu \left(\frac{1}{3} \sigma^3(\mathbf{x} \cdot \mathbf{m}) - \frac{1}{4} \sigma^2(\mathbf{x} \cdot \mathbf{m}) E[\sigma^2(\mathbf{x} \cdot \mathbf{m})] \right)$$

- Risk function (expected loss):

$$R_{\mathbf{m}}(\mathbf{x}) = -\mu \left(\frac{1}{3} E[\sigma^3(\mathbf{x} \cdot \mathbf{m})] - \frac{1}{4} E^2[\sigma^2(\mathbf{x} \cdot \mathbf{m})] \right)$$

- change \mathbf{m} to minimize $R_{\mathbf{m}}$ (gradient descent):

$$\begin{aligned} \frac{d\mathbf{m}}{dt} &= -\frac{\partial}{\partial \mathbf{m}} R_{\mathbf{m}} \\ &= \mu \left(E[\sigma^2(\mathbf{x} \cdot \mathbf{m}) \sigma'(\mathbf{x} \cdot \mathbf{m}) \mathbf{x}] - \frac{1}{4} 2E[\sigma^2(\mathbf{x} \cdot \mathbf{m})] E[2\sigma(\mathbf{x} \cdot \mathbf{m}) \sigma'(\mathbf{x} \cdot \mathbf{m}) \mathbf{x}] \right) \end{aligned}$$

$$\frac{d\mathbf{m}}{dt} = \mu E[\phi(c, \theta) \sigma'(\mathbf{x} \cdot \mathbf{m}) \mathbf{x}]$$

4.3 Network with Feedforward Inhibition

Linear Neurons

- Some definitions (for the k th neuron):

$$\begin{aligned} c_k &= \mathbf{x} \cdot \mathbf{m}_k - \eta \sum_{j \neq k} \mathbf{x} \cdot \mathbf{m}_j \\ \theta_k &\equiv E[c_k^2] \end{aligned}$$

- this is the same as **one iteration** of lateral inhibition
- Risk function:

$$R = \sum_k R_k = -\sum_k \mu \left(\frac{1}{3} E[c_k^3] - \frac{1}{4} E^2[c_k^2] \right)$$

• **weight modification**

$$\begin{aligned} \frac{d\mathbf{m}_k}{dt} &= -\frac{\partial R}{\partial \mathbf{m}_k} = -\sum_j \frac{\partial R_j}{\partial \mathbf{m}_k} \\ &= -\left(\frac{\partial R_k}{\partial \mathbf{m}_k} + \sum_{j \neq k} \frac{\partial R_j}{\partial \mathbf{m}_k} \right) \\ &= -\left(\frac{\partial R_k}{\partial \mathbf{m}_k} + \sum_{j \neq k} \frac{\partial R_j}{\partial \mathbf{m}_j} \frac{\partial \mathbf{m}_j}{\partial c_j} \frac{\partial c_j}{\partial \mathbf{m}_k} \right) \\ \frac{\partial c_k}{\partial \mathbf{m}_k} &= \frac{\partial}{\partial \mathbf{m}_j} \left(\mathbf{x} \cdot \mathbf{m}_k - \eta \sum_{j \neq k} \mathbf{x} \cdot \mathbf{m}_j \right) = \mathbf{x} \\ \frac{\partial c_k}{\partial \mathbf{m}_j} &= -\eta \mathbf{x} \\ \frac{\partial \mathbf{m}_j}{\partial c_j} &= \mathbf{x}^{-1} \end{aligned}$$

$$\boxed{\frac{d\mathbf{m}_k}{dt} = -\left(\frac{\partial R_k}{\partial \mathbf{m}_k} - \eta \sum_{j \neq k} \frac{\partial R_j}{\partial \mathbf{m}_j} \right) = \mu \left(E[\phi(c_k, \theta_k) \mathbf{x}] - \eta \sum_{j \neq k} E[\phi(c_j, \theta_j) \mathbf{x}] \right)}$$

Nonlinear Neurons

$$\begin{aligned} c_k &= \sigma \left(\mathbf{x} \cdot \mathbf{m}_k - \eta \sum_{j \neq k} \mathbf{x} \cdot \mathbf{m}_j \right) \\ \frac{\partial c_k}{\partial \mathbf{m}_k} &= \sigma'(c_k) \mathbf{x} \\ \frac{\partial c_k}{\partial \mathbf{m}_j} &= -\eta \sigma'(c_k) \mathbf{x} \end{aligned}$$

$$\boxed{\frac{d\mathbf{m}_k}{dt} = -\frac{\partial R}{\partial \mathbf{m}_k} = \mu \left(E[\phi(c_k, \theta_k) \sigma'(c_k) \mathbf{x}] - \eta \sum_{j \neq k} E[\phi(c_j, \theta_j) \sigma'(c_j) \mathbf{x}] \right)}$$

4.5 Why Feedforward Inhibition?

- we use **feedforward inhibition** because it **avoids a matrix inversion**
- **feedforward inhibition** leads to the same modification as **lateral inhibition** given the **mean field approximation**
- **Lateral Inhibition**

\mathbf{c} is now a vector of values, one for each neuron. \mathbf{M} becomes the weight matrix, etc...

$$\begin{aligned} \mathbf{c} &= \mathbf{M}\mathbf{x} + \mathbf{L}\mathbf{c} \\ &= (\mathbf{1} - \mathbf{L})^{-1} \mathbf{M}\mathbf{x} \\ &= (\mathbf{1} + \underbrace{\mathbf{L}}_{\text{monosynaptic}} + \underbrace{\mathbf{L}^2}_{\text{disynaptic}} + \underbrace{\mathbf{L}^3}_{\text{trisynaptic}} + \dots) \mathbf{M}\mathbf{x} \end{aligned}$$

- **Lateral Inhibition with mean field**(Cooper and Scofield, 1988)

$$\frac{d\mathbf{m}_k}{dt} = \mu \left(E[\phi(c_k, \theta_k)\mathbf{x}] + L_o \frac{d\bar{\mathbf{m}}}{dt} \right)$$

- **Feedforward Inhibition**

Each order is like one iteration of the lateral inhibition

$$\begin{aligned} \mathbf{c}(0) &\equiv \mathbf{M}\mathbf{x} \\ \mathbf{c}(1) &= \mathbf{c}(0) + \mathbf{L}\mathbf{c}(0) = (\mathbf{1} + \mathbf{L})\mathbf{c}(0) \\ \mathbf{c}(2) &= \mathbf{c}(0) + \mathbf{L}\mathbf{c}(1) = (\mathbf{1} + \mathbf{L} + \mathbf{L}^2)\mathbf{c}(0) \\ \mathbf{c}(3) &= \mathbf{c}(0) + \mathbf{L}\mathbf{c}(2) = (\mathbf{1} + \mathbf{L} + \mathbf{L}^2 + \mathbf{L}^3)\mathbf{c}(0) \\ &\vdots \end{aligned}$$

- **First Order Feedforward Inhibition**

$$\begin{aligned} \mathbf{c}(1) = \mathbf{c}(0) + \mathbf{L}\mathbf{c}(0) &\rightarrow c_i = \mathbf{m}_i \cdot \mathbf{x} + \sum_j L_{ij}(\mathbf{m}_i \cdot \mathbf{x}) \\ \frac{d\mathbf{m}_k}{dt} &= \mu \left(E[\phi(c_k, \theta_k)\mathbf{x}] - \sum_j L_{kj} E[\phi(c_j, \theta_j)\mathbf{x}] \right) \\ \bar{\mathbf{m}} &\equiv \frac{1}{N} \sum_j m_j \\ \left\{ \text{assume } L_{ij} = L_{ji} \Rightarrow L_o = \frac{1}{N} \sum_{ij} L_{ij} = \sum_i L_{ij} \right\} &\Rightarrow \\ N \frac{d\bar{\mathbf{m}}}{dt} &= \mu \left(\sum_k E[\phi(c_k, \theta_k)\mathbf{x}] - \sum_j L_o E[\phi(c_j, \theta_j)\mathbf{x}] \right) \\ &= \frac{1 + L_o}{L_o} \mu \sum_k \sum_j L_{kj} E[\phi(c_k, \theta_k)\mathbf{x}] \\ \Rightarrow \frac{d\mathbf{m}_k}{dt} &= \mu \left(E[\phi(c_k, \theta_k)\mathbf{x}] + \frac{L_o}{1 + L_o} \frac{d\bar{\mathbf{m}}}{dt} \right) \end{aligned}$$

- Therefore, the First Order Feedforward is the same as the mean field approximation of the lateral inhibition, with only the change in the mean field constant: $L_o \rightarrow \frac{L_o}{1+L_o}$

5 Fixed Points for Different Input Conditions

- Our goal is to determine the **fixed points** of the **Risk function**, and their **stability** for several input conditions: **linearly independant inputs**, **noise with zero mean**, and **patterned input with noise**

5.1 Linearly Independant Inputs

- Some definitions:

- N linearly independant inputs: $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$
- Probability for finding each input: $P(\mathbf{x}_i) \equiv p_i$ with $\sum_i p_i = 1$

- weight modification: $\dot{\mathbf{m}} = \epsilon\mu(t)(\mathbf{x} \cdot \mathbf{m})(\mathbf{x} \cdot \mathbf{m} - \theta) = \epsilon\mu(t)(\mathbf{x} \cdot \mathbf{m})(\mathbf{x} \cdot \mathbf{m} - E[(\mathbf{x} \cdot \mathbf{m})^2]) =$
- response vector \mathbf{v} has elements v_i equal to the response of the neuron to each input vector \mathbf{x}_i . ie.

$$\begin{pmatrix} (\cdots & \mathbf{x}_1 & \cdots) \\ (\cdots & \mathbf{x}_2 & \cdots) \\ \vdots \\ (\cdots & \mathbf{x}_N & \cdots) \end{pmatrix} \mathbf{m} = \mathbf{v}$$

- weight stationary points occur when, for all i , $\mathbf{x}_i \cdot \mathbf{m} = 0$ or $\mathbf{x}_i \cdot \mathbf{m} = \theta \neq 0$. This is the same as saying the values of the vector \mathbf{v} are either zero or θ . There are 2^N ways of achieving this, so there are 2^N stationary points.

- **Stationary Point 0:**

$$\mathbf{x}_i \cdot \mathbf{m} = 0 \quad \forall i \Rightarrow \mathbf{v}_{(0)} = (0, \dots, 0)$$

- **Stationary Point 1:**

$$\begin{aligned} \mathbf{x}_1 \cdot \mathbf{m} &= \theta \neq 0 \\ \mathbf{x}_i \cdot \mathbf{m} &= 0 \quad \forall (i > 1) \\ \theta = E[(\mathbf{x} \cdot \mathbf{m})^2] &= \sum_{i=1}^N p_i (\mathbf{x}_i \cdot \mathbf{m})^2 \\ &= p_1 (\mathbf{x}_1 \cdot \mathbf{m})^2 \\ \mathbf{x}_1 \cdot \mathbf{m} = \theta &= p_1 (\mathbf{x}_1 \cdot \mathbf{m})^2 \\ &\Rightarrow \mathbf{v}_{(1)} = \left(\frac{1}{p_1}, 0, \dots, 0 \right) \end{aligned}$$

- **Stationary Point 2:**

$$\begin{aligned} \mathbf{x}_2 \cdot \mathbf{m} &= \theta \neq 0 \\ \mathbf{x}_i \cdot \mathbf{m} &= 0 \quad \forall (i \neq 2) \\ \theta = E[(\mathbf{x} \cdot \mathbf{m})^2] &= \sum_{i=1}^N p_i (\mathbf{x}_i \cdot \mathbf{m})^2 \\ &= p_2 (\mathbf{x}_2 \cdot \mathbf{m})^2 \\ \mathbf{x}_2 \cdot \mathbf{m} = \theta &= p_2 (\mathbf{x}_2 \cdot \mathbf{m})^2 \\ &\Rightarrow \mathbf{v}_{(2)} = \left(0, \frac{1}{p_2}, 0, \dots, 0 \right) \end{aligned}$$

- **Stationary Point 3:**

$$\begin{aligned} \mathbf{x}_1 \cdot \mathbf{m} = \mathbf{x}_2 \cdot \mathbf{m} &= \theta \neq 0 \\ \mathbf{x}_i \cdot \mathbf{m} &= 0 \quad \forall (i > 2) \\ \theta = E[(\mathbf{x} \cdot \mathbf{m})^2] &= \sum_{i=1}^N p_i (\mathbf{x}_i \cdot \mathbf{m})^2 \\ &= p_1 (\mathbf{x}_1 \cdot \mathbf{m})^2 + p_2 (\mathbf{x}_2 \cdot \mathbf{m})^2 \\ \mathbf{x}_1 \cdot \mathbf{m} = \theta &= p_1 (\mathbf{x}_1 \cdot \mathbf{m})^2 + p_2 (\mathbf{x}_2 \cdot \mathbf{m})^2 \\ &\Rightarrow \mathbf{v}_{(3)} = \left(\frac{1}{p_1 + p_2}, \frac{1}{p_1 + p_2}, 0, \dots, 0 \right) \end{aligned}$$

- **Stationary Point $2^N - 1$:**

$$\mathbf{v}_{(3)} = (1, \dots, 1)$$

5.1.1 Stability of the Fixed Points

- Find the gradient of the Risk with respect to the weight vector \mathbf{m}

$$\begin{aligned} R_{\mathbf{m}}(\mathbf{x}) &= -\frac{1}{3}E[(\mathbf{x} \cdot \mathbf{m})^3] + \frac{1}{4}E^2[(\mathbf{x} \cdot \mathbf{m})^2] \\ \nabla_{\mathbf{m}} R_{\mathbf{m}}(\mathbf{x}) &= -E[(\mathbf{x} \cdot \mathbf{m})^2 \mathbf{x}] + E[(\mathbf{x} \cdot \mathbf{m})^2] E[(\mathbf{x} \cdot \mathbf{m}) \mathbf{x}] \end{aligned}$$

- Form the matrix of second derivatives of the Risk function with respect to the weight vector \mathbf{m}

$$\begin{aligned} (\mathbf{H})_{ij} &\equiv \frac{d^2 R_{\mathbf{m}}(\mathbf{x})}{dm_i dm_j} \\ \mathbf{H} &= (\nabla_{\mathbf{m}} \otimes \nabla_{\mathbf{m}}) R_{\mathbf{m}}(\mathbf{x}) \\ &= -2E[(\mathbf{x} \cdot \mathbf{m})(\mathbf{x} \otimes \mathbf{x})] + E[(\mathbf{x} \cdot \mathbf{m})^2] E[(\mathbf{x} \otimes \mathbf{x})] + 2E[(\mathbf{x} \cdot \mathbf{m}) \mathbf{x}] \otimes E[(\mathbf{x} \cdot \mathbf{m}) \mathbf{x}] \end{aligned}$$

- If \mathbf{H} is **positive definite** for \mathbf{m} equal to a stationary weight vector, then it will have only **positive eigenvalues** which corresponds to a minimum. Otherwise the stationary point is **unstable**.

- Notational reminder:

$$p\text{th fixed point: } \mathbf{v}_{(p)} = \left(\underbrace{v_1}_{(\mathbf{x}_1 \cdot \mathbf{m})}, \underbrace{v_2}_{(\mathbf{x}_2 \cdot \mathbf{m})}, \underbrace{v_3}_{(\mathbf{x}_3 \cdot \mathbf{m})}, \dots, \underbrace{v_N}_{(\mathbf{x}_N \cdot \mathbf{m})} \right)$$

- **Stationary Point 0:** $\mathbf{v}_{(0)} = (0, \dots, 0)$

$$\mathbf{H}|_{\mathbf{v}_{(0)}} = 0 \Rightarrow \text{unstable}$$

- **Stationary Point 1:** $\mathbf{v}_{(1)} = \left(\frac{1}{p_1}, 0, \dots, 0 \right)$

- Look at each term in \mathbf{H} .

$$\begin{aligned} E[(\mathbf{x} \cdot \mathbf{m})(\mathbf{x} \otimes \mathbf{x})] &= \sum_{i=1}^N p_i (\mathbf{x}_i \cdot \mathbf{m})(\mathbf{x}_i \otimes \mathbf{x}_i) \\ &= \mathbf{x}_1 \otimes \mathbf{x}_1 \\ E[(\mathbf{x} \cdot \mathbf{m})^2] &= \frac{1}{p_1} \\ E[(\mathbf{x} \otimes \mathbf{x})] &\equiv \mathbf{B} \text{ (which is positive definite)} \\ E[(\mathbf{x} \cdot \mathbf{m}) \mathbf{x}] &= \sum_{i=1}^N p_i (\mathbf{x}_i \cdot \mathbf{m}) \mathbf{x}_i \\ &= \mathbf{x}_1 \\ \mathbf{H}|_{\mathbf{v}_{(1)}} &= -2(\mathbf{x}_1 \otimes \mathbf{x}_1) + \frac{1}{p_1} \mathbf{B} + 2(\mathbf{x}_1 \otimes \mathbf{x}_1) \\ &= \frac{1}{p_1} \mathbf{B} \Rightarrow \text{stable} \end{aligned}$$

- **Stationary Point 3:** $\mathbf{v}_{(3)} = \left(\frac{1}{p_1+p_2}, \frac{1}{p_1+p_2}, 0, \dots, 0 \right)$

- Look at each term in \mathbf{H} .

$$\begin{aligned} E[(\mathbf{x} \cdot \mathbf{m})(\mathbf{x} \otimes \mathbf{x})] &= \sum_{i=1}^N p_i (\mathbf{x}_i \cdot \mathbf{m})(\mathbf{x}_i \otimes \mathbf{x}_i) \\ &= p_1 (\mathbf{x}_1 \cdot \mathbf{m})(\mathbf{x}_1 \otimes \mathbf{x}_1) + p_2 (\mathbf{x}_2 \cdot \mathbf{m})(\mathbf{x}_2 \otimes \mathbf{x}_2) \end{aligned}$$

$$\begin{aligned}
&= \frac{p_1}{p_1 + p_2} \mathbf{x}_1 \otimes \mathbf{x}_1 + \frac{p_2}{p_1 + p_2} \mathbf{x}_2 \otimes \mathbf{x}_2 \\
E[(\mathbf{x} \cdot \mathbf{m})^2] &= \frac{1}{p_1 + p_2} \\
E[(\mathbf{x} \otimes \mathbf{x})] &\equiv \mathbf{B} \\
E[(\mathbf{x} \cdot \mathbf{m})\mathbf{x}] &= \sum_{i=1}^N p_i (\mathbf{x}_i \cdot \mathbf{m}) \mathbf{x}_i \\
&= \frac{p_1}{p_1 + p_2} \mathbf{x}_1 + \frac{p_2}{p_1 + p_2} \mathbf{x}_2 \\
\mathbf{H}|_{\mathbf{v}^{(3)}} &= \left(\frac{-2p_1}{p_1 + p_2} + \frac{2p_1^2}{(p_1 + p_2)^2} \right) (\mathbf{x}_1 \otimes \mathbf{x}_1) + \left(\frac{-2p_2}{p_1 + p_2} + \frac{2p_2^2}{(p_1 + p_2)^2} \right) (\mathbf{x}_2 \otimes \mathbf{x}_2) \\
&\quad + \frac{1}{p_1 + p_2} \mathbf{B} + \frac{2p_1 p_2}{(p_1 + p_2)^2} (\mathbf{x}_1 \otimes \mathbf{x}_2 + \mathbf{x}_2 \otimes \mathbf{x}_1)
\end{aligned}$$

- Define a direction orthogonal to all but either \mathbf{x}_1 or \mathbf{x}_2 , whichever is least likely. This direction we will show to be unstable. We will choose $p_1 > p_2$, with no loss of generality and define the direction vector \mathbf{y} such that

$$\mathbf{y} \cdot \mathbf{x}_i = 0 \quad \forall i \neq 2$$

Note that this does *not* imply that \mathbf{y} is parallel to \mathbf{x}_2 , just orthogonal to all others.

- Look at each term of $\mathbf{y}^T \mathbf{H} \mathbf{y}$, which is equal to \mathbf{H} in the \mathbf{y} direction

$$\begin{aligned}
\mathbf{y}^T (\mathbf{x}_i \otimes \mathbf{x}_i) \mathbf{y} &= 0 \quad \forall i \neq 2 \\
\mathbf{y}^T (\mathbf{x}_i \otimes \mathbf{x}_2) \mathbf{y} = \mathbf{y}^T (\mathbf{x}_2 \otimes \mathbf{x}_i) \mathbf{y} &= 0 \quad \forall i \neq 2 \\
\mathbf{y}^T (\mathbf{x}_2 \otimes \mathbf{x}_2) \mathbf{y} &= (\mathbf{x}_2 \cdot \mathbf{y})^2 \\
\mathbf{y}^T \mathbf{B} \mathbf{y} &= p_2 (\mathbf{x}_2 \cdot \mathbf{y})^2 \\
\mathbf{y}^T \mathbf{H} \mathbf{y} &= \left(\frac{-2p_2}{p_1 + p_2} + \frac{2p_2^2}{(p_1 + p_2)^2} \right) (\mathbf{x}_2 \cdot \mathbf{y})^2 + \frac{p_2}{p_1 + p_2} (\mathbf{x}_2 \cdot \mathbf{y})^2 \\
&= \frac{(\mathbf{x}_2 \cdot \mathbf{y})^2}{(p_1 + p_2)^2} (p_2(p_2 - p_1)) < 0 \Rightarrow \text{unstable}
\end{aligned}$$

5.2 Noise with Zero Mean

- \mathbf{x} is white noise with zero mean

$$\begin{aligned}
E[(\mathbf{x} \cdot \mathbf{m})^3] &= 0 \\
E[(\mathbf{x} \cdot \mathbf{m})^2] &\geq 0 \\
R_{\mathbf{m}}(\mathbf{x}) &= -\mu \left(\frac{1}{3} E[(\mathbf{x} \cdot \mathbf{m})^3] - \frac{1}{4} E^2[(\mathbf{x} \cdot \mathbf{m})^2] \right) \\
&\geq 0
\end{aligned}$$

- only fixed point is $\mathbf{m} = \mathbf{0}$

5.3 Patterned Input with Noise

- \mathbf{x} is patterned input, \mathbf{d} , with small zero mean white noise, \mathbf{n}

$$\begin{aligned}
E[(\mathbf{x} \cdot \mathbf{m})^3] &= E[(\mathbf{d} \cdot \mathbf{m} + \mathbf{n} \cdot \mathbf{m})^3] \\
&= E[(\mathbf{d} \cdot \mathbf{m})^3 + 3(\mathbf{d} \cdot \mathbf{m})^2(\mathbf{n} \cdot \mathbf{m}) + 3(\mathbf{d} \cdot \mathbf{m})(\mathbf{n} \cdot \mathbf{m})^2 + (\mathbf{n} \cdot \mathbf{m})^3] \\
&= E[(\mathbf{d} \cdot \mathbf{m})^3] + 3E[(\mathbf{d} \cdot \mathbf{m})(\mathbf{n} \cdot \mathbf{m})^2]
\end{aligned}$$

$$\begin{aligned}
\{\sigma^2 \equiv E[(\mathbf{n} \cdot \mathbf{m})^2]\} &\Rightarrow \\
E[(\mathbf{x} \cdot \mathbf{m})^3] &= E[(\mathbf{d} \cdot \mathbf{m})^3] + 3\sigma^2 E[(\mathbf{d} \cdot \mathbf{m})] \\
E[(\mathbf{x} \cdot \mathbf{m})^2] &= E[(\mathbf{d} \cdot \mathbf{m} + \mathbf{n} \cdot \mathbf{m})^2] \\
&= E[(\mathbf{d} \cdot \mathbf{m})^2 + 2(\mathbf{d} \cdot \mathbf{m})(\mathbf{n} \cdot \mathbf{m}) + (\mathbf{n} \cdot \mathbf{m})^2] \\
&= E[(\mathbf{d} \cdot \mathbf{m})^2] + \sigma^2 \\
R_{\mathbf{m}}(\mathbf{x}) &= -\mu \left(\frac{1}{3} E[(\mathbf{x} \cdot \mathbf{m})^3] - \frac{1}{4} E^2[(\mathbf{x} \cdot \mathbf{m})^2] \right) \\
&= -\mu \left(\frac{1}{3} E[(\mathbf{d} \cdot \mathbf{m})^3] - \frac{1}{4} E^2[(\mathbf{d} \cdot \mathbf{m})^2] \right) \\
&\quad - \mu \sigma^2 \left(E[(\mathbf{d} \cdot \mathbf{m})] - \frac{1}{2} E[(\mathbf{d} \cdot \mathbf{m})^2] - \frac{\sigma^2}{4} \right) \\
&= R_{\mathbf{m}}(\mathbf{d}) + O(\sigma^2)
\end{aligned}$$

- The fixed points are robust to small noise

6 Deprivation Experiments

6.1 Normal Rearing

- Same as **linearly independant inputs**

6.2 Monocular Deprivation

- **inputs** and **weights** for left and right eye: $\mathbf{x} \equiv \begin{pmatrix} \mathbf{d}^l \\ \mathbf{d}^r \end{pmatrix}$, $\mathbf{m} \equiv \begin{pmatrix} \mathbf{m}^l \\ \mathbf{m}^r \end{pmatrix}$
- $\mathbf{d}^l \equiv$ white noise, $\mathbf{d}^r \equiv$ patterned input

$$\begin{aligned}
R_{\mathbf{m}}(\mathbf{x}) &= -\frac{1}{3} E[(\mathbf{x} \cdot \mathbf{m})^3] + \frac{1}{4} E^2[(\mathbf{x} \cdot \mathbf{m})^2] \\
&= -\frac{1}{3} E[(\mathbf{d}^r \cdot \mathbf{m}^r + \mathbf{d}^l \cdot \mathbf{m}^l)^3] + \frac{1}{4} E^2[(\mathbf{d}^r \cdot \mathbf{m}^r + \mathbf{d}^l \cdot \mathbf{m}^l)^2]
\end{aligned}$$

- do same algebra as above, with $v^2 \equiv E[(\mathbf{d}^l \cdot \mathbf{m}^l)^2]$

$$\begin{aligned}
R_{\mathbf{m}}(\mathbf{x}) &= \frac{1}{3} E[(\mathbf{d}^r \cdot \mathbf{m}^r)^3] - \frac{1}{4} E^2[(\mathbf{d}^r \cdot \mathbf{m}^r)^2] + v^2 \left(-E[(\mathbf{d}^r \cdot \mathbf{m}^r)] + \frac{1}{2} E[(\mathbf{d}^r \cdot \mathbf{m}^r)^2] + \frac{v^2}{4} \right) \\
&= R_{\mathbf{m}^r}(\mathbf{d}^r) + v^2 \left(-E[(\mathbf{d}^r \cdot \mathbf{m}^r)] + \frac{1}{2} E[(\mathbf{d}^r \cdot \mathbf{m}^r)^2] + \frac{v^2}{4} \right)
\end{aligned}$$

- $R_{\mathbf{m}^r}(\mathbf{d}^r)$ has just the **same fixed points** as **normal rearing**
- if $E[(\mathbf{d}^r \cdot \mathbf{m}^r)] < \frac{1}{2} E[(\mathbf{d}^r \cdot \mathbf{m}^r)^2]$, which is *easily achieved when the dimensionality is larger than 2*, then the second part of $R_{\mathbf{m}}(\mathbf{x})$ is **positive** which yields $\mathbf{m}^l = \mathbf{0}$ the only solution

6.3 Binocular Deprivation

- Same as **noise with zero mean**

6.4 Reversed Suture

- (fixed points) Same as **monocular deprivation**

A Needed formulae and theorems

•

$$g(y) = E_x[f(x, y)] \Rightarrow \frac{\partial}{\partial y} g(y) = E_x \left[\frac{\partial}{\partial y} f(x, y) \right]$$

•

$$\begin{aligned} \frac{\partial z}{\partial \mathbf{x}} \equiv \nabla_{\mathbf{x}} z &= \left(\frac{\partial z}{\partial x_1} \hat{\mathbf{x}}_1 + \frac{\partial z}{\partial x_2} \hat{\mathbf{x}}_2 + \dots \right) \\ &= \left(\frac{\partial z}{\partial t} \frac{\partial t}{\partial x_1} \hat{\mathbf{x}}_1 + \frac{\partial z}{\partial t} \frac{\partial t}{\partial x_2} \hat{\mathbf{x}}_2 + \dots \right) \\ &= \frac{\partial z}{\partial t} \left(\frac{\partial}{\partial x_1} \hat{\mathbf{x}}_1 + \frac{\partial}{\partial x_2} \hat{\mathbf{x}}_2 + \dots \right) t \end{aligned}$$

$$\boxed{\frac{\partial z}{\partial \mathbf{x}} = \frac{\partial z}{\partial t} \frac{\partial t}{\partial \mathbf{x}}}$$

•

$$\begin{aligned} \frac{\partial f(\mathbf{x} \cdot \mathbf{a})}{\partial \mathbf{x}} \equiv \nabla_{\mathbf{x}} f(\mathbf{x} \cdot \mathbf{a}) &= \left(\frac{\partial}{\partial x_1} \hat{\mathbf{x}}_1 + \frac{\partial}{\partial x_2} \hat{\mathbf{x}}_2 + \dots \right) f(\mathbf{x} \cdot \mathbf{a}) \\ &= \hat{\mathbf{x}}_1 (a_1 f'(c)|_{\mathbf{x} \cdot \mathbf{a}}) + \hat{\mathbf{x}}_2 (a_2 f'(c)|_{\mathbf{x} \cdot \mathbf{a}}) + \dots \\ \{\hat{\mathbf{a}}_i = \hat{\mathbf{x}}_i\} \Rightarrow \nabla_{\mathbf{x}} f(\mathbf{x} \cdot \mathbf{a}) &= f'(c)|_{\mathbf{x} \cdot \mathbf{a}} \mathbf{a} \end{aligned}$$

$$\boxed{\frac{\partial f(\mathbf{x} \cdot \mathbf{a})}{\partial \mathbf{x}} = f'(c)|_{\mathbf{x} \cdot \mathbf{a}} \mathbf{a}}$$

•

$$\begin{aligned} \frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x_1} \frac{\partial x_1}{\partial t} + \frac{\partial z}{\partial x_2} \frac{\partial x_2}{\partial t} + \dots \\ &= (\nabla_{\mathbf{x}} z) \cdot \frac{\partial \mathbf{x}}{\partial t} \end{aligned}$$

$$\boxed{\frac{\partial z}{\partial t} = \frac{\partial z}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}}{\partial t}}$$

- *Theorem:* if \mathbf{x}_o extremizes $f(\mathbf{x})$ on curve $\mathbf{r}(t)$, then $\nabla f(\mathbf{x}_o) \perp \mathbf{r}(t_o)$ where t_o is defined such that $\mathbf{r}(t_o) = \mathbf{x}_o$.

Proof:

$$\begin{aligned} \frac{d}{dt} f(\mathbf{r}(t)) &= \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) \\ \frac{d}{dt} f(\mathbf{r}(t)) \Big|_{t_o} = 0 &= \nabla f(\mathbf{r}(t_o)) \cdot \mathbf{r}'(t_o) \\ &\Rightarrow \nabla f(\mathbf{r}(t_o)) \perp \mathbf{r}'(t_o) \\ \frac{d}{dt} \mathbf{r}(t) &\equiv \mathbf{r}'(t) \text{ is tangent to } \mathbf{r}(t) \\ &\Rightarrow \nabla f(\mathbf{r}(t_o)) \perp \mathbf{r}(t_o) \end{aligned}$$

- *Theorem:* if \mathbf{x}_o extremizes $f(\mathbf{x})$ such that constraint $g(\mathbf{x}) = 0$, then $\nabla f(\mathbf{x}_o) \parallel \nabla g(\mathbf{x}_o)$.

Proof: **define** tangent vector $\mathbf{t}(\mathbf{x})$ such that $\nabla g(\mathbf{x}) \cdot \mathbf{t}(\mathbf{x}) = 0$.

$$\begin{aligned} (\max / \min) [f(\mathbf{x})] \text{ with } g(\mathbf{x}) = 0 &\Rightarrow (\max / \min) [f(\mathbf{x})] \text{ on curve } \mathbf{t}(\mathbf{x}) \\ &\Rightarrow \nabla f(\mathbf{x}_o) \cdot \mathbf{t}(\mathbf{x}_o) = 0 = \nabla g(\mathbf{x}_o) \cdot \mathbf{t}(\mathbf{x}_o) \\ &\Rightarrow \nabla f(\mathbf{x}_o) = \lambda \nabla g(\mathbf{x}_o) \end{aligned}$$

- To obtain the stability of the stationary points of the function $f(\mathbf{x})$:

- Form the matrix of second derivatives

$$\mathbf{H}_{ij} = \frac{d^2 f(\mathbf{x})}{dx_i dx_j} = (\nabla_{\mathbf{x}} \otimes \nabla_{\mathbf{x}}) f(\mathbf{x})$$

- Find its eigenvalues
- If they are all positive \Rightarrow minimum. all negative \Rightarrow maximum. some of each \Rightarrow saddle point.

- Outer product

$$\begin{aligned} \mathbf{x} \otimes \mathbf{y} &\equiv \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_N \\ x_2 y_1 & \ddots & & \\ \vdots & & \ddots & \\ x_N y_1 & & & x_N y_N \end{pmatrix} \\ &= \mathbf{xy}^T \end{aligned}$$

- Multiplication with outer product

$$\begin{aligned} (\mathbf{x} \otimes \mathbf{y}) \mathbf{w} &= \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_N \\ x_2 y_1 & \ddots & & \\ \vdots & & \ddots & \\ x_N y_1 & & & x_N y_N \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix} \\ &= (\mathbf{x} \cdot \mathbf{w}) \mathbf{y}^T \\ \mathbf{w}^T (\mathbf{x} \otimes \mathbf{y}) &= (w_1, \cdots, w_N) \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_N \\ x_2 y_1 & \ddots & & \\ \vdots & & \ddots & \\ x_N y_1 & & & x_N y_N \end{pmatrix} \\ &= \mathbf{x} (\mathbf{y} \cdot \mathbf{w}) \end{aligned}$$

- Outer Second Derivative

$$\begin{aligned} (\nabla_{\mathbf{x}} \otimes \nabla_{\mathbf{x}}) f(\mathbf{x} \cdot \mathbf{a}) &= \frac{d^2 f(\mathbf{x})}{dx_i dx_j} \\ &= \begin{pmatrix} (\cdots \frac{\partial}{\partial x_1} \nabla_{\mathbf{x}} f(\mathbf{x} \cdot \mathbf{a}) \cdots) \\ (\cdots \frac{\partial}{\partial x_2} \nabla_{\mathbf{x}} f(\mathbf{x} \cdot \mathbf{a}) \cdots) \\ \vdots \\ (\cdots \frac{\partial}{\partial x_N} \nabla_{\mathbf{x}} f(\mathbf{x} \cdot \mathbf{a}) \cdots) \end{pmatrix} \\ &= \begin{pmatrix} (\cdots \frac{\partial}{\partial x_1} f'(c)|_{\mathbf{x} \cdot \mathbf{a}} \mathbf{a} \cdots) \\ (\cdots \frac{\partial}{\partial x_2} f'(c)|_{\mathbf{x} \cdot \mathbf{a}} \mathbf{a} \cdots) \\ \vdots \\ (\cdots \frac{\partial}{\partial x_N} f'(c)|_{\mathbf{x} \cdot \mathbf{a}} \mathbf{a} \cdots) \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&= f''(c)|_{\mathbf{x}\cdot\mathbf{a}} \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} (a_1, \dots, a_N) \\
&= f''(c)|_{\mathbf{x}\cdot\mathbf{a}} (\mathbf{a} \otimes \mathbf{a})
\end{aligned}$$

- *Theorem:* if the N vectors \mathbf{x}_i are **linearly independent** and **span the space**, then the matrix $\mathbf{B} \equiv E[\mathbf{x} \otimes \mathbf{x}]$ is **positive definite**.

Proof: \mathbf{B} is **positive definite** if

$$\mathbf{q}^T \mathbf{B} \mathbf{q} > 0 \quad \forall \mathbf{q} \neq \mathbf{0}$$

$$\begin{aligned}
(\mathbf{B})_{ij} &= \sum_k p_k (\mathbf{x}_k)_i (\mathbf{x}_k)_j \\
(\mathbf{q}^T \mathbf{B} \mathbf{q}) &= \sum_{ijk} p_k q_i q_j (\mathbf{x}_k)_i (\mathbf{x}_k)_j \\
&= \sum_k p_k (\mathbf{x}_k \cdot \mathbf{q}) (\mathbf{x}_k \cdot \mathbf{q}) \\
&= \sum_k p_k (\mathbf{x}_k \cdot \mathbf{q})^2 > 0
\end{aligned}$$

References

- Cooper, L. N. and Scofield, C. L. (1988). Mean-field theory of a neural network. *Proceedings of the National Academy of Science*, 85:1973–1977.
- Freeman, R., Mallach, R., and Hartley, S. (1981). Responsivity of normal kitten striate cortex deteriorates after brief binocular deprivation. *Journal of Neurophysiology*, 45(6):1074–1084.
- Freeman, R. and Olson, C. (1982). Brief periods of monocular deprivation in kittens: Effects of delay prior to physiological study. *Journal of Neurophysiology*, 47(2):139–150.
- Intrator, N. and Cooper, L. N. (1992). Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5:3–17.
- Mioche, L. and Singer, W. (1989). Chronic recordings from single sites of kitten striate cortex during experience-dependent modifications of receptive-field properties. *J. Neurophysiol*, 62:85–197.