

**The Role of the Environment in Synaptic Plasticity:
Towards an Understanding of Learning and Memory**

by

Brian S. Blais

B.A. Wesleyan University 1992

M.Sc Brown University 1994

Thesis

Submitted in partial fulfillment of the requirements for the
Degree of Doctor of Philosophy in the Department of
Physics at Brown University.

May 1998

© Copyright 1998
by
Brian S. Blais

This dissertation by **Brian S. Blais**
is accepted in its present form by the Department of Physics
as satisfying the dissertation requirements for the degree of
Doctor of Philosophy.

Date _____
Prof. Leon N Cooper

Recommended to the Graduate Council

Date _____
Prof. J. Valles

Date _____
Prof. R. Pelcovits

Approved by the Graduate Council

Date _____

Vita

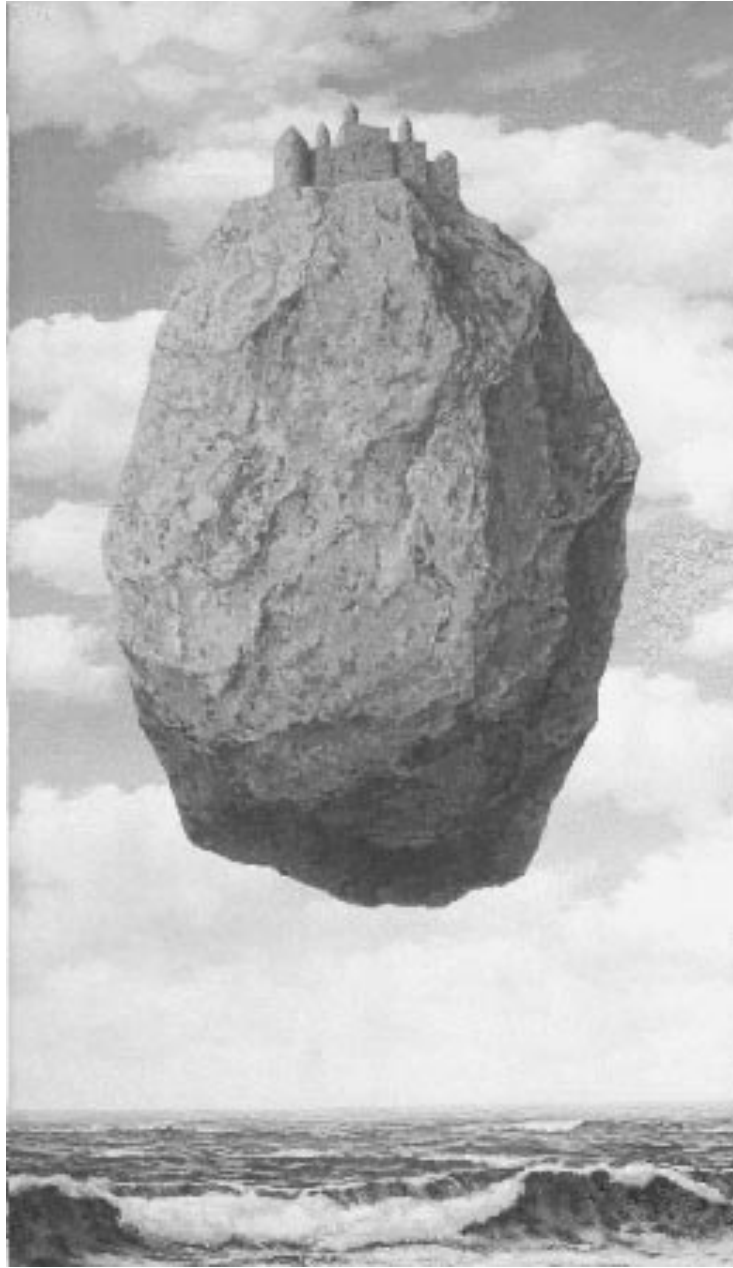
Born August 17, 1970, Vermont

Education

- B.A. Wesleyan University 1992
- M.Sc Brown University 1994

Publications

- **Blais B.S.**, Intrator N., Shouval H., and Cooper L. N. 1998. Receptive field formation in natural scene environments: comparison of single cell learning rules. *Submitted to Neural Computation*.
- **Blais B.S.** 1997. Physicist Studying the Brain. *The Catalyst*.
- **Blais B.S.** , Shouval H., and Cooper L. N. 1997. Dynamics of synaptic plasticity: A comparison between models and experimental results in visual cortex. In *The Neurobiology of Computation: Proceedings of the fifth CNS conference* .
- Perrone, M.P., and **B.S. Blais**. 1995. Regression NSS: An Alternative to Cross Validation. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory*. 385-391.
- Carroll, D.L., D.L. Doering, and **B.S. Blais**. 1992. Dynamic Redistribution of Excess Charge During Photoemission in an Electron Bombarded Glass-Ceramic. *In Proceedings of the Materials Research Society*. 235:395.
- Carroll, D.L., D.L. Doering, and **B.S. Blais**. 1992. Thermally- and Optically-Stimulated Exo-electron Emission from an Electron Beam Irradiated Glass-Ceramic Material. *Journal of Vacuum Science Technology*. 10(7):2863.



They are a mystery, and I am both terrified and reassured that there are still wonders in the Universe...that we haven't explained everything.

G'Kar

Acknowledgments

There are many people who need to be thanked, without whom I could have never accomplished this work. I'd like to thank my advisor, Professor Leon N Cooper for reminding me of the important questions and keeping me from getting lost in the details of computer simulations. His ability to see through the details of any complex problem in order to pull out the important aspects has been a constant guide for me in my work. I also want to thank my readers, Professor Jim Valles and Professor Bob Pelcovits, both of whom have been a wonderful support for me throughout graduate school.

I'd like to thank all of those people in the lab who have dealt with all of my quirks through the years. This includes most recently Ann Lee, Pedja Neskovic, Omer Artun, and David Goldberg. I would also like to thank Roger Blumberg and Wes Wallace, for providing some of the more philosophical aspects to the lab. I thank Mike Perrone, for bringing me into this work when I was a first year graduate student, Charlie Law for providing the starting point of much this work, and Nathan Intrator for his insights and support. In particular, I'd like to thank Harel Shouval. He has been incredibly supportive for the entire time I have worked in the lab, and has been a ready reference on almost anything. I could not have done this work without his support and guidance.

I would like to thank all of those people who have provided a good example for me. More specifically, I'd like to thank Ross Larsen, Joe Adams, and Dave Carroll for constantly reminding me what good physics is all about. I'd like to thank Joe Barry (my high school physics teacher), Professor Richard Lindquist (one of my college professors), and, again, Professor Leon N Cooper for being exemplary teachers and inspiring my own teaching experiences. I'd like to thank Andrew Kayser, who was willing to come out of the trenches and keep me honest.

Finally, I'd like to thank my family for all of their support. I would like especially to thank my fiancée Heidi Schattle, who has been a great support. I love her dearly.

Contents

Vita	iii
Acknowledgments	v
1 Introduction	1
1.1 Motivation and Key Assumptions	3
1.1.1 Characteristics of Neurons	3
1.1.2 Activation Equation	3
1.1.3 Selectivity	6
1.1.4 Modification Equations	6
1.2 One Dimensional Model: BCM	9
1.2.1 Fixed Point	9
1.2.2 Oscillations	12
1.2.3 Conclusions about 1D BCM	13
1.3 Some Properties of the PCA learning rule	15
1.4 One Dimensional Model: PCA	16
1.5 Two Dimensional Model: BCM	16
1.5.1 Fixed Points	17
1.6 Two Dimensional Model: PCA	18
1.7 LTP and LTD	19
1.7.1 Discussion	21
2 Visual Deprivation	24
2.1 Introduction	24
2.2 The Visual System	24
2.3 Experimentally Modifying the Input Environment	29

2.3.1	Normal Development	29
2.3.2	Binocular Deprivation	31
2.3.3	Monocular Deprivation and Reverse Suture	31
2.3.4	Strabismus	32
2.3.5	Stripe and Strobe Rearing	32
2.3.6	Conclusions about Visual Environment Modification	33
2.4	Modeling the Development Orientation Selectivity and Ocular Dominance	33
2.4.1	Normal Rearing (NR)	33
2.4.2	Deprivation	35
2.5	Time Course of Deprivation: an exploration of parameter space	37
2.5.1	PCA: Parameter Dependence	40
2.5.2	PCA: Summary of parameter dependence	41
2.5.3	BCM: Robustness to Parameters	42
2.5.4	BCM: Dependence on τ	43
2.5.5	BCM: Dependence on the noise	44
2.5.6	BCM: Finding a valid parameter regime	45
2.5.7	BCM: Conclusions about parameter dependence	46
2.6	Summary of Comparison between BCM and PCA	47
2.7	Other Models of Orientation Selectivity and Ocular Dominance	48
2.7.1	A Toy Model	50
2.7.2	Correlation-based Model of Orientation Selectivity	51
2.7.3	Correlation-based Model of Monocular Deprivation	53
2.7.4	Problems with Correlation-based Models	53
2.8	Conclusions	55
3	Projection Pursuit	57
3.1	Introduction	57
3.2	BCM and PCA	58
3.2.1	Example with two dimensional model	59
3.3	Output Distribution	60
3.4	Other Cost Functions	61
3.5	The Effect of Noise on Monocular Deprivation	65
3.5.1	Experimental Verification	66

3.5.2	Binocular Deprivation and Reverse Suture	68
3.6	A Simpler Environment	69
3.6.1	One Dimensional Example: Laplace	69
3.6.2	Two Dimensions	72
3.6.3	Monocular Deprivation	75
3.6.4	Binocular Deprivation	79
3.6.5	Reverse Suture	83
3.6.6	Strabismus	83
3.6.7	Conclusions about the Simple Environment	85
4	Extensions to the Model	87
4.1	Introduction	87
4.2	X and Y Cells	87
4.3	Direction Selectivity	89
4.4	Structure Removal: Sensitivity to Outliers	92
5	Conclusions	100
A	Some Mathematical Results	108
A.1	Discrete Versions of BCM Equations	108
A.2	Calculation and Self Consistency Check for BCM Oscillations	109
A.3	Full Solution for 2D BCM fixed points	111
A.4	Full Solution for 2D PCA fixed points	112
A.5	Stability of BCM fixed points	113
A.5.1	Some useful formulae and theorems	113
A.5.2	Defining the Cost Function	116
A.5.3	Single Nonlinear Neuron	117
A.5.4	Fixed points for N linearly independent inputs	118
A.5.5	Calculating the Stability	119
A.6	Stability of PCA fixed points	121
A.7	Deriving the Wyatt Equation	122
A.8	Classical Rearing Conditions: PCA Analysis	123
A.8.1	Normal Rearing (NR)	125
A.8.2	Monocular Deprivation (MD)	126

A.8.3	Binocular Deprivation (BD)	127
A.8.4	Reverse Suture (RS)	127
A.9	Some Useful Results From the Simple Environment	128
A.9.1	From Papoulis (1984)	128
B	Some Computational Issues	129
B.1	Code Segments for Section 2.7	129
B.1.1	MATLAB	129
B.2	Code Segments for Section 3.6	133
B.2.1	Maple	133
B.2.2	MATLAB	136

List of Figures

- 1.1 The Anatomy of a Neuron (adapted from Bear, Connors, Paradiso, *Neuroscience: Exploring the Brain*). Outgoing signals are sent along the **axon**, and incoming signals are collected in the **soma** from the **dendrites**. Communication from one neuron to another takes place across the **synapse**, where the axon from one neuron meets the dendrites of another. 3
- 1.2 Integration of Input Signals in the Soma (adapted from Bear, Connors, Paradiso, *Neuroscience: Exploring the Brain*). **(a)** A presynaptic action potential causes the postsynaptic potential, V_m , to rise in the soma. **(b)** Presynaptic signals from multiple sources have an integrated effect on the target soma. If the postsynaptic potential is high enough, the target soma will generate its own action potentials. **(c)** Integration in the soma can occur with signals coming in rapid succession. 5
- 1.3 Model Neuron. Inputs are given by $\mathbf{x} \equiv (x_1, \dots, x_n)$, weights by $\mathbf{w} \equiv (w_1, \dots, w_n)$ and output by y 6
- 1.4 Effects of the parameters on the development of the one dimensional neuron. The value of the weight, w , and the threshold, θ , are shown as functions of time for different values of the learning rate, η , and the memory constant, τ . The value of the input here is $x = 2$, so the fixed point should be $w = 1/2$ and $\theta = 1$ 11
- 1.5 Frequency of oscillations, ω , (left) and the decay constant, g , (right) as a function of the input, taken from simulation. The parameters for the top simulations are $\tau = 2000, \eta = 0.001$, whereas the parameters for the bottom simulations are $\tau = 4000, \eta = 0.0001$. The three regions denote convergence (A), damped oscillations (B), and unstable behavior (C). Also shown, with a dashed line, is the solution from Equations 1.15 and 1.16. . . . 14

1.6	LTP and LTD (results from Kirkwood and Bear (1995)). Very low frequency “baseline” stimulus is presented alternately to two independent pathways, A and B. Measurements of excitatory postsynaptic potentials (EPSPs, or simply, the activity of the cell) are performed. After high frequency stimulation (HFS) to pathway A, the response of that pathway is enhanced (LTP) and independent pathways, B, are unaffected. Low frequency stimulation (LFS) to pathway A causes a <i>reduced</i> response (LTD) in that pathway, leaving independent pathways unaffected. These forms of LTP and LTD are often called <i>homosynaptic</i> LTP/LTD, referring to the fact that only the stimulated pathway is affected.	20
1.7	Measuring the BCM $\phi(\cdot)$ function. Shown is the change in EPSP (excitatory postsynaptic potential, e.g. activity) as a function of the input frequency. Shown above is from Dudek and Bear (1992), measured in hippocampus. The total input into the cell is kept constant, for all input frequencies, so the change in activity (or synaptic efficacy) is a direct measure of the modification function, $\phi(\cdot)$. Shown below is Kirkwood et. al. (1996). The two graphs correspond to the same measurement performed on rats with no visual experience (dark reared) and those with normal visual experience (normally reared). An activity dependent shift is observed, which is consistent with the motion of the modification threshold, θ .	22
2.1	A horizontal slab through the brain exposing the visual pathway. (adapted from: Bear, Connors, Paradiso, <i>Neuroscience: Exploring the Brain</i>)	25
2.2	Example Receptive Field for an ON-center LGN cell. Shown are spots of light on different parts of the receptive field (left) and the resulting spike trains (right). Shining a spot of light in the center yields a strong response (top). A spot of light in the surround gives an inhibited response (middle), and outside of the receptive field, or at least beyond the surround, gives spontaneous activity (bottom). Copyright 1995, Izumi Ohzawa.	26
2.3	Reverse Correlation Example. Random stimulus is presented, and those patterns which give responses are added up to yield the receptive field.	27
2.4	Example Receptive Fields for cells in the LGN and visual cortex obtained using reverse correlation. Copyright 1995-1997, Izumi Ohzawa.	28
2.5	Possible construction of an oriented receptive field by the convergence of several LGN cells, with center-surround receptive fields (proposed by Hubel and Wiesel to explain orientation selective cortical cells). Shown is a layer of LGN cells projecting to a single cortical cell. If there are strong connections between LGN cells falling in a line, then a bar stimulus along that orientation will excite the centers of many LGN cells, and cause a larger response in the cortical cell.	30

2.6	Model Architecture. Shown are the image plane (top), the left and right retinal cells, the left and right LGN cells, and the single cortical cell (bottom). In the image plane are drawn sample center-surround receptive fields, for the retinal cells highlighted. Nearby retinal cells see nearby points in the image plane. Retinal cells project directly to LGN cells, on a one-to-one basis. A circular patch of LGN cells projects to the single cortical cell. These projections form the input vector, \mathbf{x} , for which there is a corresponding weight vector, \mathbf{w} . The output of the cell is given simply as $y = \sigma(\mathbf{w} \cdot \mathbf{x})$	34
2.7	Input Environment. Shown are the original images (top) and the retinally processed images used as the actual inputs to the neuron (bottom). The images are processed with a difference of Gaussians (DOG) filter (center, right), which is used as a model of the receptive field properties of the retinal cells (center left).	35
2.8	Example Receptive Fields from BCM and PCA trained in a natural scene environment. Shown are the left and right eye receptive fields for BCM (top) and PCA (bottom). . .	36
2.9	Example PCA Simulations. Left: Final weight configuration. Right: Maximum response to oriented stimuli, as a function of time. Simulations from top to bottom are as follows. Normal Rearing (NR): both eyes presented with patterned input. Monocular Deprivation (MD): following NR, one eye is presented with noisy input and the other with patterned input. Reverse Suture: following MD, the eye given noisy input is now given patterned input, and the other eye is given noisy input. Binocular Deprivation (BD): following NR, both eyes are given noisy input. It is important to note that for PCA if Binocular Deprivation is run longer, selectivity will not be lost.	37
2.10	Example BCM Simulations. Left: Final weight configuration. Right: Maximum response to oriented stimuli, as a function of time. Simulations from top to bottom are as follows. Normal Rearing (NR): both eyes presented with patterned input. Monocular Deprivation (MD): following NR, one eye is presented with noisy input and the other with patterned input. Reverse Suture: following MD, the eye given noisy input is now given patterned input, and the other eye is given noisy input. Binocular Deprivation (BD): following NR, both eyes are given noisy input. It is important to note that for BCM if Binocular Deprivation is run longer, selectivity will eventually be lost.	38
2.11	Example of a Response Half-Time Measurement. This is an illustration of the procedure for measuring the half-times. Though the example uses a BCM simulation, the specific numbers are not important. The time is measured for the neuron to either rise or fall, half way between its minimum and maximum responses.	39
2.12	The dependence of the half-fall times, $\mathcal{T}_{\text{fall}}^{\text{MD}}$, $\mathcal{T}_{\text{fall}}^{\text{BD}}$, and $\mathcal{T}_{\text{fall}}^{\text{RS}}$ on the memory constant τ	44
2.13	The dependence of the half-fall times, $\mathcal{T}_{\text{fall}}^{\text{MD}}$, $\mathcal{T}_{\text{fall}}^{\text{BD}}$, and $\mathcal{T}_{\text{fall}}^{\text{RS}}$ on the standard deviation of the noise, σ . ($\eta = 5e - 6$, $\tau = 1000$).	44

-
- 2.14 Schematic for finding a valid parameter regime. Areas in the parameter space are found which yield consistent half-time ratios, $\mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$, and $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$. The time ratios involving $\mathcal{T}_{\text{rise}}^{\text{RS}}$ are only used as a possible consistency check, and not to determine the valid range, because of problems discussed earlier. The memory constant, τ , has no appreciable effect on the half-times so it is not used to determine the valid parameter regime. 45
- 2.15 The dependence of the half-fall time ratios $\mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$, and $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ on the standard deviation of the noise, σ . ($\eta = 5e - 6$, $\tau = 1000$). The valid ranges, from experiment, are shown with the arrows. 46
- 2.16 The dependence of the half-fall time ratios $\mathcal{T}_{\text{fall}}^{\text{BD}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$, and $\mathcal{T}_{\text{fall}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{MD}}$ on the learning rate, η . ($\sigma = 0.8$, $\tau = 1000$). The valid ranges, from experiment, are shown with the arrows. 46
- 2.17 The dependence of the half-fall time ratios $\mathcal{T}_{\text{rise}}^{\text{RS}}/\mathcal{T}_{\text{fall}}^{\text{BD}}$, and $\mathcal{T}_{\text{rise}}^{\text{RS}}\mathcal{T}_{\text{fall}}^{\text{MD}}$ on the standard deviation of the noise, σ . ($\eta = 5e - 6$, $\tau = 1000$). The valid ranges, from experiment, are shown with the arrows. 47
- 2.18 Toy model correlation functions and eigenvectors. Shown is the toy correlation function, $\mathbf{C}(x - x') = 1 + q \cos(\pi(x - x'))$, as a one dimensional function (upper left, both **A** and **B**) and as a two dimensional function (upper right, both **A** and **B**). The three eigenvectors are shown in the lower plots of **A** and **B**, with the corresponding eigenvalues, λ . Plots in **A** are for $q = 1$, where the DC solution is dominant. Plots in **B** are for $q = 4$, where the AC solution is dominant. Small deviations from the expected values of λ , and the small non-DC parts of the DC solution are merely numerical in origin, and should be ignored. 52
- 2.19 Correlation-based model for orientation selectivity. Shown are the sum and difference correlation functions, \mathbf{C}^{S} and \mathbf{C}^{D} respectively, for a model of ON and OFF channel inputs, as a function of the distance (upper and lower left plots), and in a matrix form (upper and lower center graphs) where the four dimensions have been collapsed onto two. The banded structure comes both from the arbor function and the correlation function restricting the extent of the receptive field. The four receptive fields on the upper right are the dominant eigenvectors of \mathbf{C}^{S} with labeled eigenvalues. The four receptive fields on the lower right are the dominant eigenvectors of \mathbf{C}^{D} , also with labeled eigenvalues. . . 54
- 3.1 Example directions of significant structure. Directions of high kurtosis (left) and multimodality (right) are two examples of “interesting” directions in the space. 58
-

-
- 3.2 Example Cost Maximization in 2D. Shown (above) are sample input patterns, \mathbf{x}_1 and \mathbf{x}_2 , chosen so both the solutions for BCM and for PCA, $\mathbf{w}_i^{\text{BCM}}$ (circles) and $\mathbf{w}_i^{\text{PCA}}$ (squares), fall on a unit circle. Also shown (below) are the BCM cost function (left) and PCA cost function (right) as a function of angle around that unit circle. The angles where both sets of solutions fall are labeled on both cost function graphs. It is clear that the BCM cost function is maximized when the weight is equal to a solution of the original BCM equation (Equation 1.4). The different weight vectors which maximize the PCA cost are the solutions of the original PCA equation (Equation 1.3). 60
- 3.3 Output Distribution From Natural Scenes. Shown are an example receptive field (left) trained with BCM, the output distribution of a neuron with this receptive field on a linear scale (center) and a log scale (right). The distribution is achieved by performing a histogram of the output values (pre-sigmoid), over the entire environment. Also shown, with a dotted line, is a Gaussian distribution with the same variance. 61
- 3.4 Example learning rules, as a function of the output of the cell. Shown are Quadratic BCM, Kurtosis 1 and 2, and Skewness 1 and 2. 66
- 3.5 The half-time for the loss of response to the closed eye in monocular deprivation, as a function of the closed eye noise level (structured input variance=1). The half-times are scaled so that the half-time is set to 1 at a noise level of unit variance, $\sigma = 1$. The rules in Class 1 (QBCM, K_1 , and S_1) are shown above, and the rules in Class 2 (non-linear PCA, K_2 , and S_2) are shown below. The Class 1 learning rules have a *faster* loss of response to the closed eye, for *more noise* into the closed eye. The Class 2 learning rules have the opposite behavior, though much less pronounced. 67
- 3.6 Odd low dimensional behavior of BCM. Shown is a sample taken from a 1D-2 eye input distribution where both eyes receive the same input, corrupted by noise. Input points are shown as small dots, the BCM fixed points are shown with asterisks and their direction shown with an arrow for clarity. All of the input falls along the (left eye)=(right eye) line. On the left is shown a situation where the neuron can find no selective fixed points (i.e. those fixed points where *most* responses are around zero, and a few responses have significant non-zero value). In these cases, the fixed points found (shown in asterisks) are non-physiological: the cell becomes responsive to an odd combination of left and right eyes. The one to the right shows a case where the neuron becomes selective, and the fixed point is physiological: both eyes respond equally. 73
-

3.7	2D Deprivation. Normal rearing (A) is modeled with numbers chosen from a Laplace distribution presented to each eye. The eyes see identical inputs, so the input distribution (samples shown with dots) lies along the 45° line. The initial weights are small, and the final weights are equal for each eye (Equation 3.57 for BCM and Equation 3.58 for kurtosis). Monocular deprivation (B) is modeled with Laplace numbers to one eye, representing the structure from the natural environment, and uniform (or Gaussian) numbers to the other eye, representing the noise of activity from the closed eye. The initial weight is from normal rearing, and the final weight is in the direction of the open eye: the cell comes to respond only to the open eye. Reverse suture (C) is modeled with Laplace numbers to one eye and uniform (or Gaussian) numbers to the other eye, following monocular deprivation. Binocular deprivation (D) is modeled using uniform (or Gaussian) noise presented to both eyes, following normal rearing.	74
3.8	Binocular Deprivation Fixed Points in a low dimensional environment. For the Gaussian case (left), the fixed points for both K_2 and R_{BCM} fall on a circle. For the uniform case (right), they point in the direction of the corners of the distribution.	82
3.9	Shown are the unrectified (left) and rectified (right) Gaussian distributions, along with the rectifying sigmoid. In the high noise case (above) the sigmoid introduces a significant skew, whereas in the low noise case (below) the sigmoid doesn't change the distribution much.	83
3.10	2D Strabismus. Strabismus is modeled with numbers chosen from a Laplace distribution presented to each eye. The eyes have <i>independent</i> inputs (samples shown with dots). The initial weight is from normal rearing, and the final weight is in the direction of one of the eyes alone: the cell comes to respond only to only one eye.	84
4.1	Size of X and Y cell receptive fields, as measured by Linsenmeier et. al. (1982), and fit to center-surround difference of Gaussians. (From Figure 8, and Paragraph 2 on Page 1179 in Linsenmeier). The mean center and surround values found are $c = 0.83, s = 3.2$ for X cells and $c = 2.9, s = 4.3$ for Y cells.	88
4.2	Filters and filtered images for X and Y cells. Shown is the original image (left), and the difference of Gaussian (DOG) filters for X and Y cells (above center and right). The center and surround values used are $c = 0.83, s = 3.2$ for X cells and $c = 2.9, s = 4.3$ for Y cells. Shown also are the images resulting from the application of these filters (below center and right). The X cells respond to much higher spatial frequencies in the images than the Y cells.	90
4.3	Example receptive fields trained cortical cells, with X cell input (above) and Y cell input (below).	90

4.4	Fitting a receptive field to a Gabor filter. Shown is an example receptive field (upper left), the best fit sine grating (upper right), the best fit Gaussian window (lower left), and the product of the sine grating and the Gaussian window.	91
4.5	Spatial frequency of trained cortical neurons, as a function of the center and surround sizes of retinal receptive fields.	91
4.6	Spatiotemporal contour plots of a one dimensional ST separable receptive field (left) and ST inseparable receptive field (right). A temporal cross section made vertically through the contour plot represents the temporal response function to flashed stimulus at a point x . Example response functions, for points A and B, are shown on the right of each contour plot.	95
4.7	Spatiotemporal receptive field plots of separable (top) and inseparable (bottom) receptive fields, superimposed over spatiotemporal representations of drifting sinusoidal gratings. Gratings appear “oriented” because they shift in a particular spatial direction (x axis) as they move through time (y axis). The response of the cell is the product of the stimulus with the receptive field, summed over both space and time. Thus, for inseparable receptive fields, the stimulus at the preferred space-time “orientation”, i.e. moving the preferred direction of drift, is a more effective stimulus than one moving in the non-preferred direction.	96
4.8	Example receptive fields and their orientation tuning, for a velocity of 2 pixels per iteration. The response of the cell, for a particular orientation of sine grating, is given by the radial component of the polar plots. The orientation tuning was obtained using drifting oriented sine gratings. Orientations larger than 180 degrees denotes motion in the opposite direction. Tuning curves which have a larger response for one direction than another are for direction selective cells. The PCA neuron is the only one which did not achieve direction selectivity.	97
4.9	Example receptive fields with polar tuning plots (above) for BCM, for several eye drift velocities. The direction selectivity index as a function of velocity (below) for four different learning rules. The PCA learning rule did not develop direction selectivity. The other rules show some velocity tuning: they all lose direction selectivity for either velocities which are too high or too low. The LGN lagged cells had a constant 1 iteration lag. . .	97
4.10	Patterns which yield high responses of a model neuron. The example receptive field is shown on the left. Some of the patterns which yield the strongest 1/2 percent of responses are labeled on the image on the right. These patterns are primarily the high contrast edges.	98

- 4.11 Receptive fields resulting from structure removal using the Quadratic BCM rule, the rule maximizing the multiplicative form of kurtosis and skewness. The RF on the far left for each rule was obtained in the normal input environment. The next RF to the right was obtained in a reduced input environment, whose patterns were deleted that yielded the strongest 1% of responses from the RF to the left. This process was continued for each RF from left to right, yielding a final removal of about five percent of the input patterns. 98
- 4.12 Normalized difference between RFs as a function of the percentage deleted in structure removal. The RFs were normalized, and mean zero, in order to neglect magnitude and additive constant changes. The maximum possible value of the difference is 1. 99

Chapter 1

Introduction

It is widely believed that much of learning, memory storage, and resulting organization of many parts of the brain occur due to the modification of the efficacy or strength of at least some of the synaptic junctions between neurons, thus altering the relation between one neuron and another. It is evident that it would be highly inefficient for the genetic code to specify the efficacy of each synapse in the brain. A simple alternative is for the genetic code to specify some general mechanisms for modifying the synapses, and have the environment itself supply the rest of the information. Though it is not necessary that the mechanisms behind the modification operate in exactly the same manner in all portions of the nervous system, or in all animals, we assume that there are some fundamental similarities. **The central theme of this work is the role of the environment in the development, and maintenance, of orientation selectivity and ocular dominance in the visual cortex.** Along the way, we compare statistically and biologically motivated learning rules used to model the learning in the visual cortex. I hope to show, given the current state of the experimental data, what can be nailed down theoretically and what cannot. In other words, I want give a motivation for further experiments to test *specific* aspects of the theory in order to increase our understanding of learning and memory.

The general approach is to start simply, introducing the most obvious properties of the neuron. Only after comparisons with either experiment or other learning rules force us to add modifications, will we be able to add complications to the models. Simplicity is necessary if we are to gain any insight into the problem, but it is a danger as well. Simplifying assumptions can bypass some of the important characteristics of the brain. We will focus our attention on neurons in the visual cortex, because of the vast amount of experimental data on that area. It is our hope, however, that the knowledge we gain from the study of the visual system will generalize to other parts of the brain, and thus give us an understanding about learning and memory *in general*.

Among the difficulties we face while modeling the visual cortex are

1. lack of knowledge of what the actual input signals to the visual cortical cells are
2. the appropriate rule for synaptic modification

3. adequate representation of the visual environment, which is related to Item 1
4. an adequate representation of the complex architecture of the visual cortex

Later in this chapter we discuss the assumptions made about the inputs to the visual cortical cells (Item 1) as well as introduce some of the synaptic modification rules, or learning rules (Item 2). We choose a **visual environment composed of natural scenes** (Item 3), which is an improvement over many models which use more abstract inputs. Throughout this work we explore **single cell properties** because they are the simplest to consider, and it is a first step towards understanding the more complicated problem of networks of neurons (Item 4). This work consists primarily of comparisons between different forms of the BCM(Bienenstock et al., 1982) learning rule and other single cell rules, both on experimental and theoretical grounds.

In this chapter, we introduce the properties of neurons, a description of two learning rules, the BCM and PCA rules. We will see that the properties of many proposed learning rules can be understood from variations of these two. In this chapter we point out some of their properties, including the key assumptions used. In the process we introduce the central neuronal property with which this work is based: the development of selectivity. We then explore one and two dimensional example environments, which gives us an intuition about the learning rules.

The second chapter introduces visual deprivation as a way of comparing the theory to experiment in a more realistic environment, made of natural scenes. This environment is an improvement over many models which use more abstract inputs(Erwin and Miller, 1995; Miller, 1994; Linsker, 1986; Clothiaux et al., 1991). We continue our comparison between BCM and PCA, but in more detail, allowing us to discover the necessary aspects of the learning rules to be consistent with experiment. We then explore some other models of deprivation, which have been proposed, and show how they compare to the BCM and PCA rules.

The third chapter introduces projection pursuit, which provides a theoretical framework to discuss some of the development of selectivity, and the results from visual deprivation. We present several new learning rules, which can be placed into one of two classes: a simplification which allows us to use the results from the previous chapter. We also introduce a simplified environment which has many of the properties of the natural scene environment, but permits us to do analysis.

In the fourth chapter, we introduce some extensions to the model. These extensions do not fit well into the theme of deprivation, but provide additional insight into the workings of some of the learning rules, and gives us a broader range of experimental data to which we can compare the theory. The extensions also give a glimpse of work yet to be done, and some of the potential of the models themselves.

1.1 Motivation and Key Assumptions

1.1.1 Characteristics of Neurons

Neurons communicate with one another using electrical signals called action potentials. These signals take the form of electrical potential spikes which are produced in the soma, and travel down the length of the axon (see Figure 1.1). The spikes then traverse the connection between one neuron and a target neuron at a gap called the synapse using a complicated electro-chemical mechanism. The synapses usually occur on the dendrites of the target cell, but can occur on the main body as well. The signals from all of the dendritic branches are integrated in the soma raising the postsynaptic potential, shown in Figure 1.2. If the potential is raised high enough, then the soma will then generate a new signal down its own axon, in the form of an action potential.

The efficacy of the synapse can provide a way for a cell to modify its effective input signal. If the efficacy of a synapse is low, then it will take a much stronger input signal to elicit the same response. Modification of the synaptic efficacy can then be a very convenient mechanism for attaining desired responses from a neuron in a particular input environment. When we speak of a learning rule, or of a memory storage rule, we generally are referring to synaptic modification. The specific form of the modification equation will determine, for a particular input environment, to what the neuron will become responsive or unresponsive.

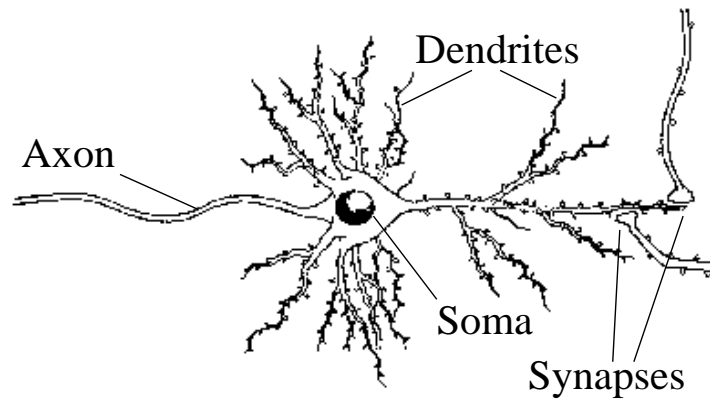


Figure 1.1: The Anatomy of a Neuron (adapted from Bear, Connors, Paradiso, *Neuroscience: Exploring the Brain*). Outgoing signals are sent along the **axon**, and incoming signals are collected in the **soma** from the **dendrites**. Communication from one neuron to another takes place across the **synapse**, where the axon from one neuron meets the dendrites of another.

1.1.2 Activation Equation

We simplify the description of the neural dynamics by choosing as variables not the instantaneous incoming time sequence of spikes in each dendrite, the instantaneous membrane potential, or the time sequence of outgoing spikes; but rather moving averages of these variables over appropriate time scales.

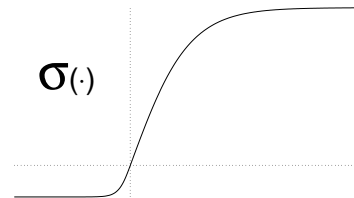
Thus, the model neuron performs a spatial integration (it integrates over all the dendrites), rather than spatiotemporal integration of the variables: the output at time t is a function only of the input at time t and the neuron parameters (synaptic efficacies, internal thresholds, etc.) at time t , independent of past history (see Figure 1.2 for examples of spatial and temporal integration). This is not to imply that the neuron has no memory, because the values of the synaptic efficacies and internal thresholds contain the “memory”, not the output of the cell itself. The time scales over which the instantaneous variables are averaged, to yield the model variables, are not explicitly stated but can possibly be inferred from the comparison of the model to experiment. This information can then be used to infer the physical processes responsible for the behavior.

The model neuron consists of a vector of inputs, \mathbf{x} , a vector of synaptic efficacies or weights, \mathbf{w} , and a scalar output, y (Figure 1.3). The values of the inputs represent an averaged presynaptic activity originating from another cell’s axon. The output is given by

$$y = \mathbf{w} \cdot \mathbf{x} \text{ (Activation equation)} \quad (1.1)$$

and represents an averaged postsynaptic activity. Neurons have a nonzero spontaneous activity when there is no input, so we may interpret the variables y and \mathbf{x} as activity *above spontaneous*. It is then reasonable to have negative values of these variables, though we may wish to limit the values of the negative activities to be consistent with this interpretation: a cell cannot have activity as far below spontaneous as it can above. This can be easily achieved by making y a non-linear, asymmetric squashing function or sigmoid of $\mathbf{w} \cdot \mathbf{x}$, where the lowest negative value of the sigmoid is smaller, in magnitude, than the highest positive value as shown here.

$y = \sigma(\mathbf{w} \cdot \mathbf{x})$, where $\sigma(\cdot)$ looks like



It is often convenient, though not completely necessary, to also allow negative weights. We can think of the weight as representing an *effective synapse*, one which could in reality be made up of multiple excitatory and inhibitory synapses or a network effect.

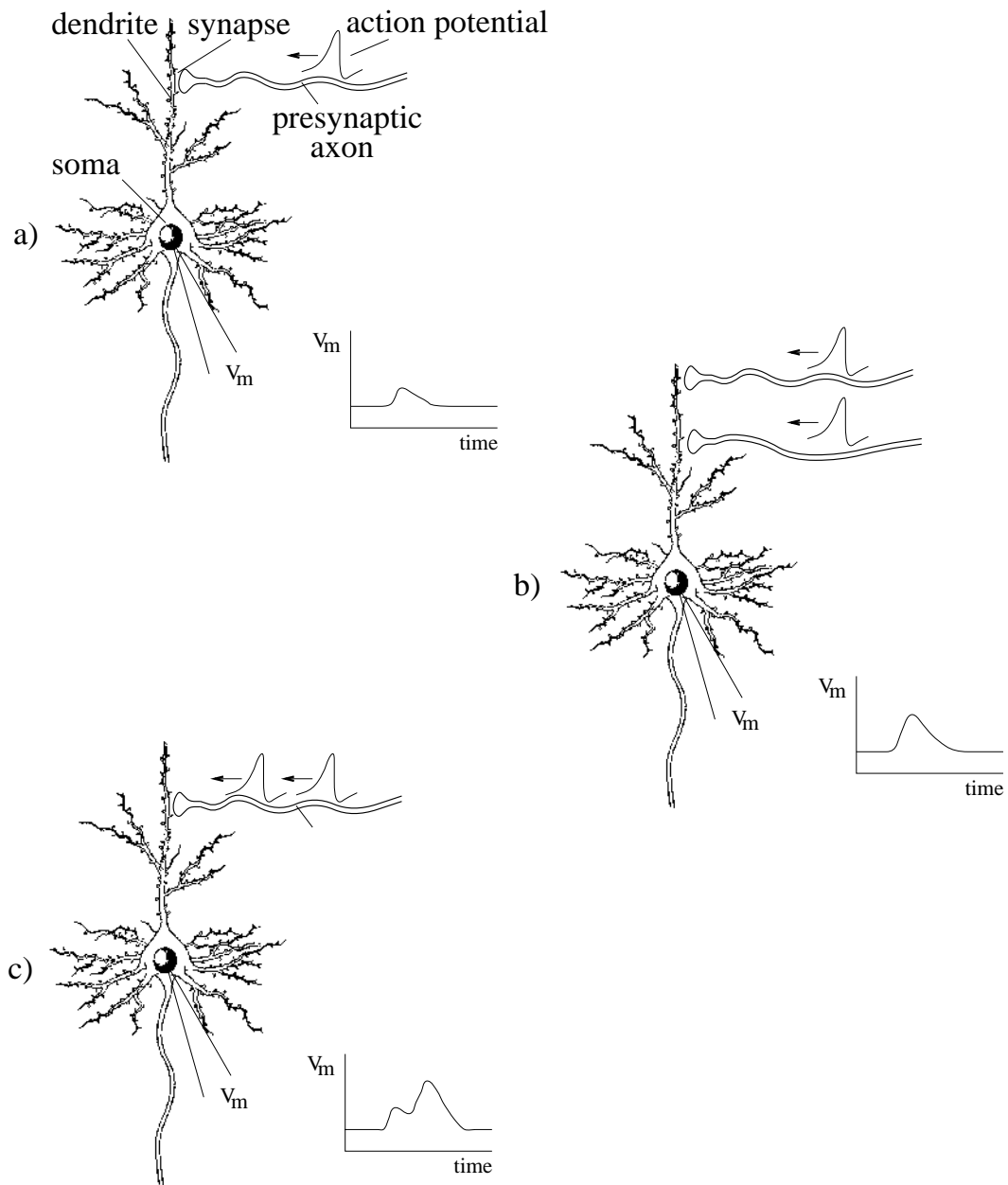


Figure 1.2: Integration of Input Signals in the Soma (adapted from Bear, Connors, Paradiso, *Neuroscience: Exploring the Brain*). **(a)** A presynaptic action potential causes the postsynaptic potential, V_m , to rise in the soma. **(b)** Presynaptic signals from multiple sources have an integrated effect on the target soma. If the postsynaptic potential is high enough, the target soma will generate its own action potentials. **(c)** Integration in the soma can occur with signals coming in rapid succession.

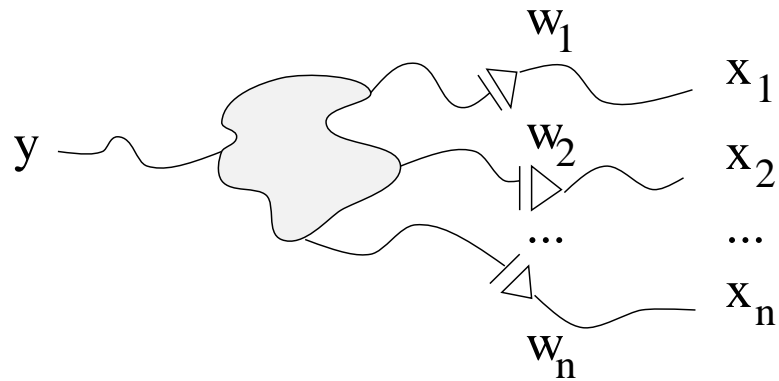


Figure 1.3: Model Neuron. Inputs are given by $\mathbf{x} \equiv (x_1, \dots, x_n)$, weights by $\mathbf{w} \equiv (w_1, \dots, w_n)$ and output by y .

1.1.3 Selectivity

It has been known for some time that sensory neurons at practically all levels display various forms of input selectivity. They may respond preferentially to a tone of a given frequency, a light spot of a given color, a light bar of a certain orientation, etc. Since the work of Hubel and Wiesel (Hubel and Wiesel, 1962) it has been known that most neurons in normally reared adult cat striate cortex (a part of the visual cortex) show a strong preference for contours of a particular orientation. Although some orientation selectivity exists in striate cortex prior to visual experience, maturation to adult levels of specificity and responsiveness requires *normal contour vision* during the first 2 months of life (for review see Fregnac and Imbert, 1984).

We will try to understand the development of selectivity, or more specifically orientation selectivity in the visual cortex. We present several synaptic modification rules, under some straightforward restrictions, and follow through with their theoretical predictions. This will then allow us to refine the models, compare them more directly with experiment, and attempt to understand some of the underlying mechanisms of learning.

1.1.4 Modification Equations

When we consider learning rules, we usually restrict ourselves to those rules which are *local* or *quasi-local*. Changes in the synapses of the neuron can only depend on information accessible to it. *Local* information refers to anything directly at the synapse. This includes the value of the synaptic efficacy and the input signal impinging on the synapse. *Quasi-local* refers to whole cell information which can be communicated to every synapse. This includes the output of the cell itself and any whole-cell thresholds. A synapse does *not* have access to the inputs along other synapses, the activities of other neurons, global information about the input patterns being presented to the neuron, etc.

The restriction to local or quasi-local rules for synapse modification is not merely for simplicity. When one writes down a learning rule, one needs to ask both *how* the information could be transmitted

to the synapse, and on what *time scale* would it need to be communicated. For example, the output of the cell is an electrical quantity, so it is likely that some kind of backward electrical signal could communicate this quantity to the synapses. Simple functions of the output of the cell could be calculated by mechanisms internal to the cell, and communicated to the synapses at a slower rate, perhaps through gene expression or conductance changes.

This way of thinking can rule out certain formulations of synaptic modification, or at least make them highly unlikely. For example, the sum of the weights is sometimes (incorrectly) considered a quasi-local property (MacKay and Miller, 1994). This has two main problems. First, in order to calculate the sum of the weights, the cell must be able to “read” the values of the weights. In practice, one only knows the value of a weight by presenting a particular input, and seeing what the output of the cell is. If the weight is stored structurally somehow, it is difficult to imagine how the cell could obtain its value. The second problem is with the time scale. In order to have a modification rule which uses the sum of the weights, the value of the sum must be communicated on the shortest time scale of modification, which is milliseconds. The cell would have to be able to collect the values of all the weights (which could be physically quite distant), sum them, and communicate that value to all of the other weights, all on the scale of milliseconds. We leave it up to the one proposing such a rule to provide a possible mechanism. The idea of locality will come up again, in Section 2.7, when we consider correlation based models.

The cornerstone of all learning rules is the Hebb rule (Hebb, 1949). The original Hebb rule states how synapse efficacies are strengthened:

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

One simple way to express this mathematically is

$$\dot{\mathbf{w}} = y\mathbf{x} \tag{1.2}$$

It is fairly clear that, in order to actually use Hebb’s principle, one must state conditions for synaptic decrease as specific as those for synaptic increase: if synapses are allowed only to increase, all synapses will eventually saturate; no information will be stored and no selectivity will develop. Some variants of the Hebb rule stabilize the weight growth through the explicit normalization of the weights. We have mentioned the problems with this, so we will not consider it further. One variant of the Hebb rule, proposed by Oja (Oja, 1982), avoids this problem. It includes a decay term on the weights, which has the effect of normalizing the weights but uses *local* information, in the form of the synapse value itself, and *quasi-local* information, in the form of the output activity:

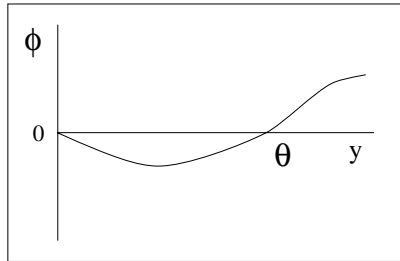
$$\dot{\mathbf{w}} = y\mathbf{x} - y^2\mathbf{w} \tag{1.3}$$

According to this rule, the increase in strength of certain synapses is accompanied by the decrease in strength of other synapses onto the same neuron. Thus, there occurs a *spatial competition between inputs*. This rule is also called the PCA rule, due to its connection to the statistical method Principle Component Analysis.

A different form of stabilization is used in the rule proposed by Bienenstock, Cooper, and Munro (Bienenstock et al., 1982), known as the BCM rule.

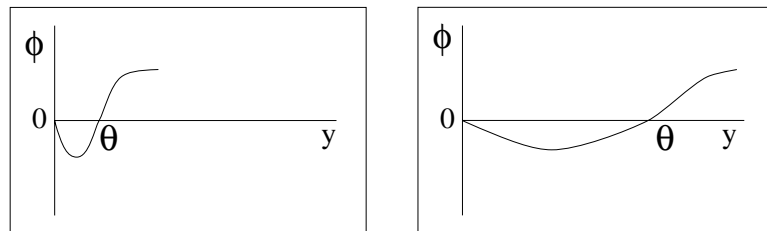
$$\dot{\mathbf{w}} = \phi(y, \theta) \mathbf{x} \tag{1.4}$$

where $\phi(y)$ is a scalar function of the postsynaptic activity, y , that changes sign at a value, θ , of the output called the modification threshold.



The vector \mathbf{w} is driven in the direction of the input \mathbf{x} if the output is large (above θ), or opposite to the direction of the input if the output is small (below θ). We may regard this as a form of *temporal competition between incoming patterns*.

If the threshold, θ , is constant then we have the same instability problem that occurred with the Hebb rule: the neuron could respond to all neurons above the threshold, causing an increase in the weights in the direction of these patterns, which leads to still stronger activity. Instead, we take θ to be a super-linear function of the output of the neuron, so it *slides* with the activity of the neuron, as shown below



The threshold acts as a negative feedback system, so the super-linearity is necessary for stability, otherwise θ would never catch up to the growth in the activity. There are many possible mathematical forms of the $\phi(\cdot)$ function. For analytic tractability we often use the quadratic form of the BCM modification equation (Intrator and Cooper, 1992):

$$\dot{\mathbf{w}} = \eta y (y - \theta) \mathbf{x} \tag{1.5}$$

$$\theta = E_{\tau} [y^2] \tag{1.6}$$

where $\phi(\cdot)$ is given simply by a parabola, with a learning rate, η , and θ is a windowed time average of the squared activity, $E_\tau [y^2]$. The windowed average has one free parameter, τ , which sets the window length and thus the time scale over which threshold averages. If $\tau \rightarrow 0$, then θ instantly adjusts to the new input, resulting in physically unreasonable fluctuations. If $\tau \rightarrow \infty$, then θ doesn't move and we are left with the stability problems noted earlier. A determination of this time scale, by comparison of the model to experiment, is therefore crucial to understanding the underlying mechanism behind the moving threshold. There are many forms we could use for the windowed average, $E_\tau[\cdot]$. For analytic convenience we commonly use an exponentially weighted windowing function.

$$\theta = \frac{1}{\tau} \int_{-\infty}^t e^{-(t-t')/\tau} y^2(t') dt' \quad (1.7)$$

Because the PCA rule, the BCM rule, and many other Hebb-based rules are stabilized versions of Hebb's postulate, it is sometimes assumed that these rules behave in similar ways. Some of the differences between the rules are quite striking, and others are more subtle. In this work we explore several lines of comparison, from theoretical to experimental, and in the process highlight the important questions in the field.

We continue now with some implementations of these rules for simple systems, to gain a better understanding of their fundamental properties. We are restricting ourselves to these two rules, at present, even though there are many more rules based on Hebb's original statement which have been used to model the visual cortex. Many of these rules, however, have the same basic properties of either the PCA rule or the BCM rule, so it is instructive to look at the simple versions of these two rules first, and then introduce more of the variants later.

1.2 One Dimensional Model: BCM

1.2.1 Fixed Point

In the one dimensional model, there is one input, x , and one weight, w , and the output of the neuron is given simply by $y = wx$. Here x is treated as a random variable, so the BCM equations become stochastic equations. These equations are

$$y = wx \quad (1.8)$$

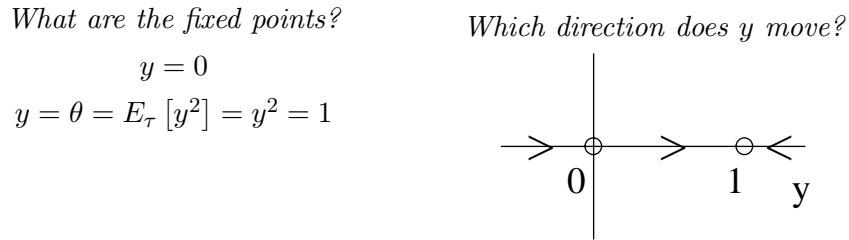
$$\dot{w} = \eta y(y - \theta)x \quad (1.9)$$

$$\theta = E_\tau [y^2] \quad (1.10)$$

The fixed point occurs when $E[\dot{w}] = 0$, where the expectation is taken over x .

In this section, we restrict ourselves to *constant* input x . We consider a less restricted case in Chapter 3. Even with this extremely restricted input structure, we obtain some remarkably complex behavior. Some of the qualitative features of this behavior, however, can give us some understanding of

more realistic models. From Equation 1.9 it is clear that the fixed points occur when $y = 0$ and when $y = \theta$. The following diagram demonstrates that, for the one dimensional case, only the $y = \theta = 1$ fixed point is stable.



The fixed point for the one dimensional, constant input, model the weight becomes

$$\begin{aligned}
 y = \theta = E_\tau [y^2] &= y^2 \\
 \Rightarrow y &= 1 = wx \\
 w &= \frac{1}{x}
 \end{aligned}
 \tag{1.11}$$

One outcome of this fixed point is that the neuron adjusts the weight to maintain a particular output activity level. If the input is lower, then the weight increases. If the input is high, then the weight decreases to compensate. There is some evidence for such a regulatory process, often called homeostasis, in experiment (Turrigiano et al., 1998). We discuss this more in Section 1.7.1.

At this point it is instructive to look at some simulation results. Although the BCM neuron attains this fixed point, a look at the dynamics getting to the fixed point will show some important properties of the BCM neuron. In order to do simulations we discretize the Equations 1.8-1.10, and iterate the resulting equations on a computer. These discrete equations (see Appendix A.1 for derivation) are

$$\begin{aligned}
 y_n &= w_n x_n \\
 \theta_{n+1} &= \theta_n + \frac{1}{\tau} (y_n^2 - \theta_n) \\
 w_{n+1} &= w_n + \eta y_n (y_n - \theta_{n+1}) x_n
 \end{aligned}
 \tag{1.12}$$

where n is the iteration number. The three parameters in the model are the input value, x , the learning rate, η , and the threshold averaging constant (or “memory constant”), τ . The behavior of the neuron will depend critically on these three parameters. Experimentally, however, we can only vary the input, x , and measure the output. Figure 1.4 are examples of the effect of these parameters on the development of the neuron.

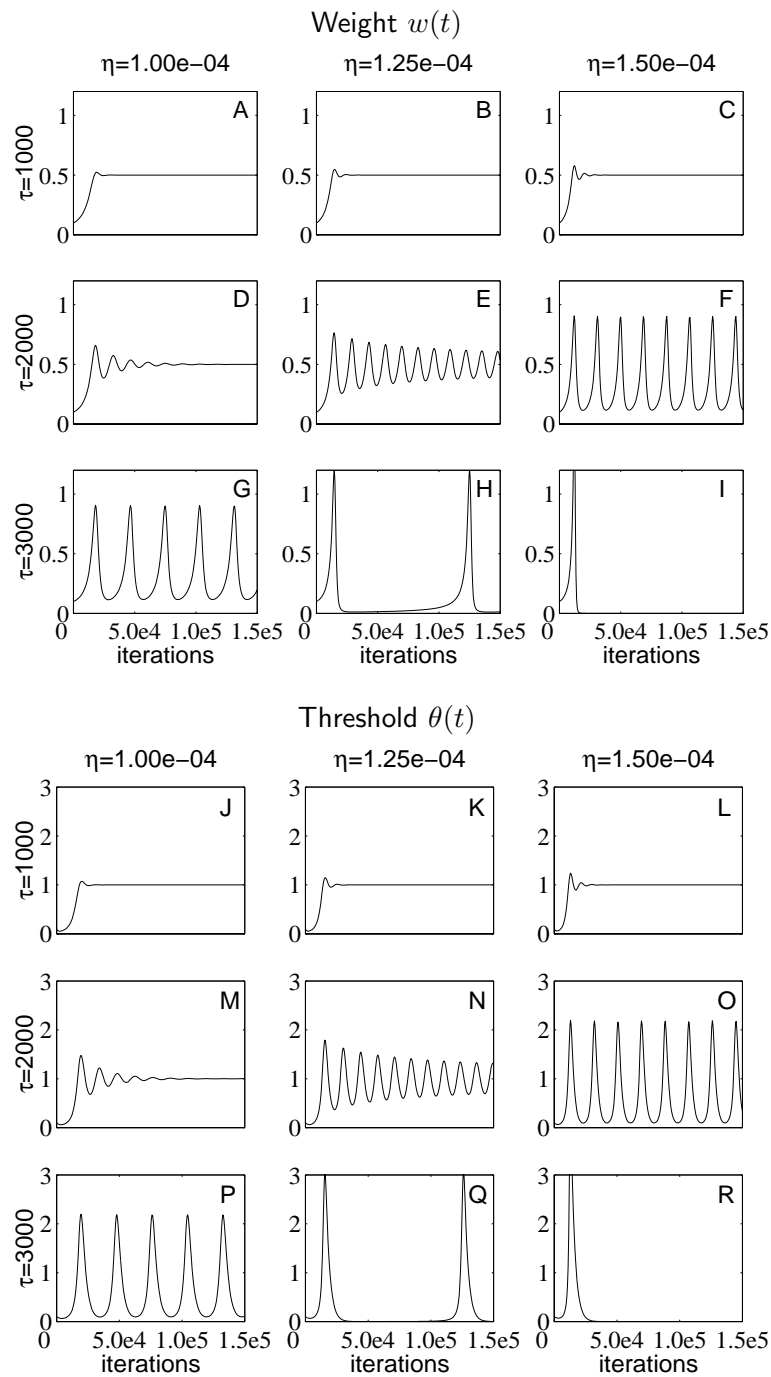


Figure 1.4: Effects of the parameters on the development of the one dimensional neuron. The value of the weight, w , and the threshold, θ , are shown as functions of time for different values of the learning rate, η , and the memory constant, τ . The value of the input here is $x = 2$, so the fixed point should be $w = 1/2$ and $\theta = 1$.

1.2.2 Oscillations

Figure 1.4A shows how the BCM neuron converges to $w = \frac{1}{2}$ and $\theta = 1$ for an input value of $x = 2$. The memory constant, τ , determines the speed of the moving threshold. If it is too large, relative to the learning rate, then the threshold moves slowly compared to the change in weight. The weight will then overshoot the fixed point until the threshold “catches up”, at which point the value of the threshold becomes larger than the output and the weights will decrease. The overshooting occurs again as the weight decreases until the threshold catches up again. This behavior can cause some mild oscillations (Figure 1.4D,E) which damp out, or it can cause (for large enough τ) some wildly non-linear oscillations (Figure 1.4G,H). To understand this, we add a damped oscillatory perturbation to the constant w_o solution given in Equation 1.11.

We have to, of course, modify the equation for θ so as not to include an integral from negative infinity.

We start with

$$\begin{aligned} y &= wx \\ \dot{w} &= \eta y(y - \theta)x \\ \theta &= \frac{1}{\tau} \int_0^t y^2(t') e^{-(t-t')/\tau} dt' + e^{-t/\tau} \theta_o \end{aligned}$$

and then assume

$$w = w_o + w_1 e^{i\omega t} e^{-gt}$$

where $w_1 \ll 1$, $w_o = 1/x$ and g is a positive damping constant.

After some nasty algebra (full details in Section A.2) we arrive at

$$\omega = \pm \frac{1}{2} \frac{\sqrt{6\tau\eta x^2 - 1 - \tau^2\eta^2 x^4}}{\tau} \quad (1.13)$$

$$g = \frac{1}{2} \frac{(1 - \tau\eta x^2)}{\tau} \quad (1.14)$$

Equations 1.13 and 1.14 suggest that the important aspects of the theory can be described using two parameters: τ and $\tau\eta x^2$ (which I will define as α). In these new parameters, Equations 1.13 and 1.14 look like:

$$\begin{aligned} \alpha &\equiv \tau\eta x^2 \\ \omega &= \pm \frac{1}{2} \frac{\sqrt{6\alpha - 1 - \alpha^2}}{\tau} \end{aligned} \quad (1.15)$$

$$g = \frac{1}{2} \frac{(1 - \alpha)}{\tau} \quad (1.16)$$

This simple appearance of the equations lets us divide the parameter space into several regions, based on the sign of ω and g . The simulations give us nice interpretations for these regions. The

comparison between the simulations and the analytical solutions is shown in Figure 1.5. The results are extremely robust to changes in the parameters.

- **Region A:** $6\alpha - 1 - \alpha^2 \leq 0$

In this region the frequency is either zero or pure imaginary. Simulations show that this results in the convergence of the neuron with no oscillations, like Figure 1.4A,B. Therefore when $\alpha \leq 3 - 2\sqrt{2}$ we get convergence.

- **Region B:** $6\alpha - 1 - \alpha^2 > 0$ and $\alpha < 1$ ($g > 0$)

In this region we get simple, damped oscillations, like Figure 1.4D,E. The simulations and Equations 1.15 and 1.16 are in good agreement in this region. Note that, as we increase the input into the neuron, the frequency of oscillations increases and the rate of damping *decreases*.

- **Region C:** $\alpha > 1$ ($g < 0$)

In this region we expect our theory to break down. If the damping constant is negative, then we get exponentially growing solutions, and our long-time approximation falls apart. Simulations show that the oscillations become increasingly nonlinear as α gets larger, like Figure 1.4H,I.

1.2.3 Conclusions about 1D BCM

It is clear that the stability of the BCM neuron is dependent most strongly on three parameters. Two of them, the learning rate η and the memory constant τ , are intrinsic to the neuron. The third, the input value x , comes from the environment. Though we have not implemented it here, it could be possible that the neuron adjusts its intrinsic parameters to fit the particular environment. This could be done mostly beforehand, by genetics, because neurons in particular parts of the brain can expect to see particular types of inputs.

Experimentally, we may be able measure the intrinsic properties, like η and τ , by careful measurements of the oscillation frequency and damping constant. One can easily solve Equations 1.16 and 1.15 for τ in terms of g and ω .

$$\begin{aligned} (g^2 + \omega^2) \tau^2 + 2g\tau - 1 &= 0 \\ \tau &= \frac{-g + \sqrt{2g^2 + \omega^2}}{g^2 + \omega^2} \end{aligned} \quad (1.17)$$

Since this expression for the memory constant does not depend on the value of the input, it could be valuable in experimentally determining this parameter. If one knows the input, then one can use Equation 1.16 to determine the learning rate, η , as well.

In the next chapter I determine through an alternate route that the value of τ is on the order of 1-30 minutes. This places an upper bound on the frequency of oscillations of $\frac{1}{\tau} \approx 0.017\text{Hz}$. Though

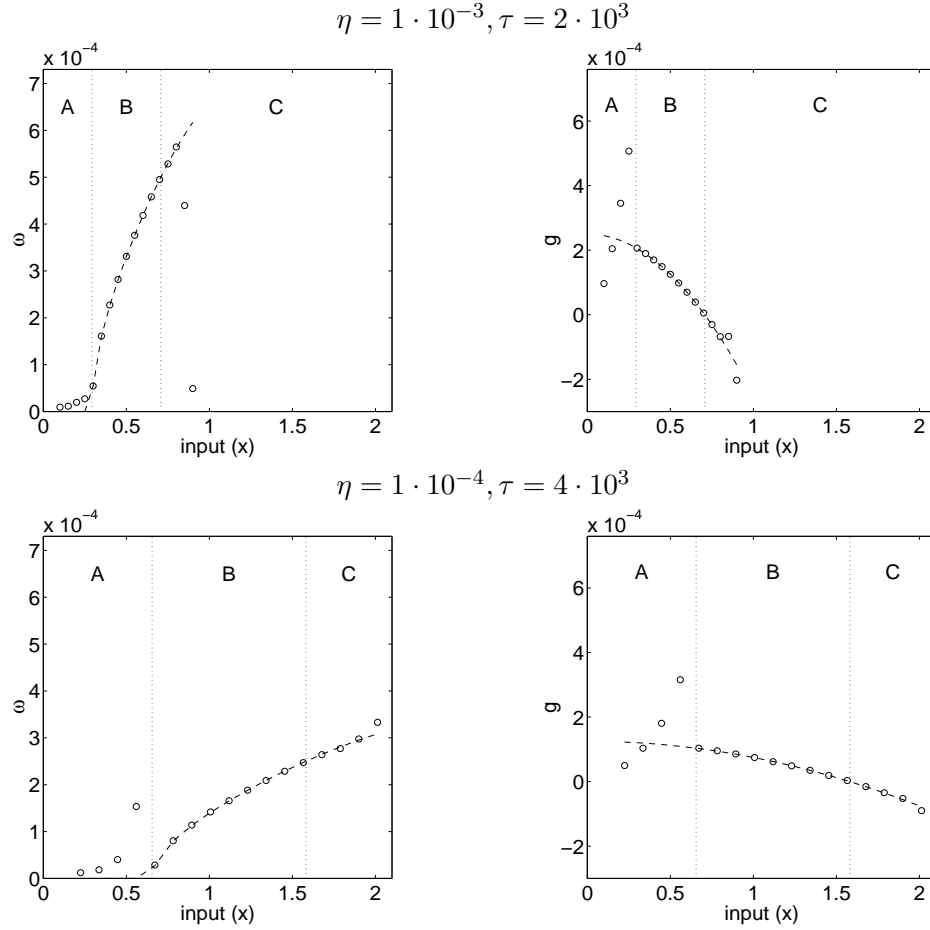


Figure 1.5: Frequency of oscillations, ω , (left) and the decay constant, g , (right) as a function of the input, taken from simulation. The parameters for the top simulations are $\tau = 2000, \eta = 0.001$, whereas the parameters for the bottom simulations are $\tau = 4000, \eta = 0.0001$. The three regions denote convergence (A), damped oscillations (B), and unstable behavior (C). Also shown, with a dashed line, is the solution from Equations 1.15 and 1.16.

many people have reported oscillations in the brain (Crick and Koch, 1990; Brown et al., 1996), and some have implicated them in types of processing including synchrony and occlusion continuity, there is no reason to think that the oscillations discussed here serve a functional role or are the same kind. For instance, the oscillations reported commonly in the literature occur on a fairly fast time scale (≈ 100 Hz), which rules out synaptic plasticity as a mechanism. The oscillations presented earlier are due to changes in synaptic efficacy; changes which occur on much longer time scales than 0.01 seconds. Therefore, if we want to measure the BCM oscillations, we should be able to distinguish them quite easily from other types of oscillations. Measurements precise enough to deal with these problems are, for the most part, only now starting to be done, so we may need to wait for a few years to get answers to these questions.

1.3 Some Properties of the PCA learning rule

We start with the equations for Oja's stabilized Hebb rule, or the so-called PCA rule

$$\begin{aligned} y &= \mathbf{w} \cdot \mathbf{x} \\ \dot{\mathbf{w}} &= y\mathbf{x} - y^2\mathbf{w} \end{aligned} \quad (1.3)$$

We then look for the fixed points

$$\begin{aligned} E[\dot{\mathbf{w}}] &= 0 & (1.18) \\ &= E[\mathbf{x}(\mathbf{x} \cdot \mathbf{w}) - (\mathbf{x} \cdot \mathbf{w})^2\mathbf{w}] \\ &= E[(\mathbf{x}\mathbf{x}^T)\mathbf{w} - (\mathbf{w}^T\mathbf{x}\mathbf{x}^T\mathbf{w})\mathbf{w}] \\ &\equiv \mathbf{C}\mathbf{w} - (\mathbf{w}^T\mathbf{C}\mathbf{w})\mathbf{w} \end{aligned} \quad (1.19)$$

where the correlation matrix is defined $\mathbf{C} \equiv E[\mathbf{x}\mathbf{x}^T]$. Notice that $\mathbf{w}^T\mathbf{C}\mathbf{w}$ is just a number, so the solution for the fixed points becomes the solution of the eigenvalue equation

$$\mathbf{C}\mathbf{w} = \lambda\mathbf{w} \quad (1.20)$$

It follows that the weight at the fixed-points are normalized.

$$\lambda = \mathbf{w}^T\mathbf{C}\mathbf{w} = \mathbf{w}^T\lambda\mathbf{w} = \lambda|\mathbf{w}|^2 \quad (1.21)$$

$$\Rightarrow |\mathbf{w}|^2 = 1 \quad (1.22)$$

It also follows, from stability analysis (see Section A.6), that the weights become parallel to the eigenvector of \mathbf{C} with the *maximum eigenvalue*. This direction is the direction of the *maximum variance* in the data, and is called the principle component.

1.4 One Dimensional Model: PCA

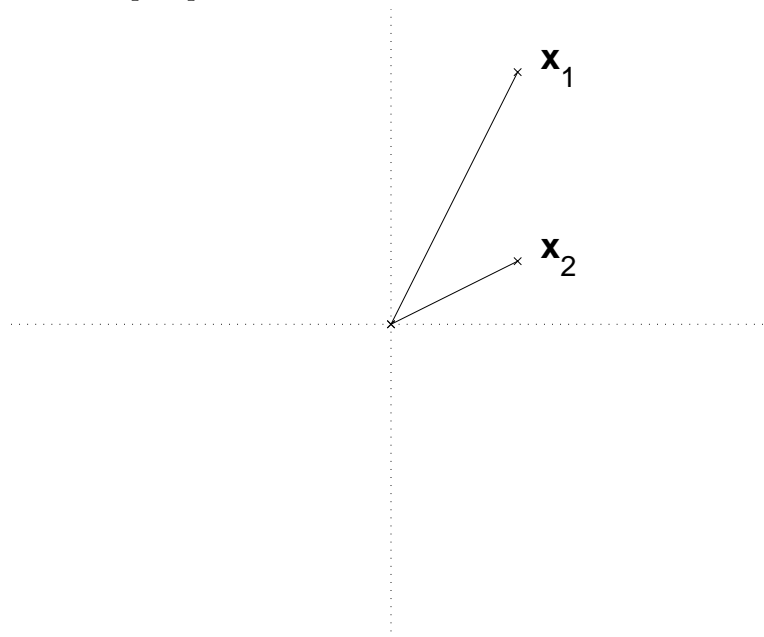
Using the same one dimensional model introduced in Section 1.2, and using Wyatt's (Wyatt and Elfadel, 1995) full dynamical solution for PCA shown in Appendix A.7, we obtain the following equation for the weight dynamics

$$w = \frac{e^{x^2 t} w_o}{(e^{2x^2 t} w_o^2 + 1 - w_o^2)^{1/2}} \quad (1.23)$$

The dynamics in this case are trivial: an exponential convergence to the value $w = 1$, where the input value x determines the rate. There are already some qualitative differences between these two rules, in this simplified 1-D case. As one increases the input to the neuron, we predict from the PCA rule that the convergence will be *faster*, whereas (from Equation 1.16) we would predict from the BCM rule that the convergence will be *slower*. This difference is amenable to experiment, and could provide a way to distinguish between possible learning rules. We will see that this difference also works its way into the higher dimensional environments, and yields qualitative differences which are more easily compared to biological systems.

1.5 Two Dimensional Model: BCM

Although the one dimensional model gives us an idea about the dynamics, one cannot obtain selectivity with this restricted environment. We therefore take the next easiest step and introduce a two dimensional model. A more thorough investigation of the multi-dimensional case, including a full stability analysis, is shown in Appendix A.5. In this model we will restrict ourselves to two input patterns, \mathbf{x}_1 and \mathbf{x}_2 , presented to the neuron with equal probabilities.



The BCM equations we use are

$$y = \mathbf{w} \cdot \mathbf{x} \quad (1.24)$$

$$\dot{\mathbf{w}} = \eta y(y - \theta)\mathbf{x} \quad (1.25)$$

$$\theta = E_{\tau} [y^2] \quad (1.26)$$

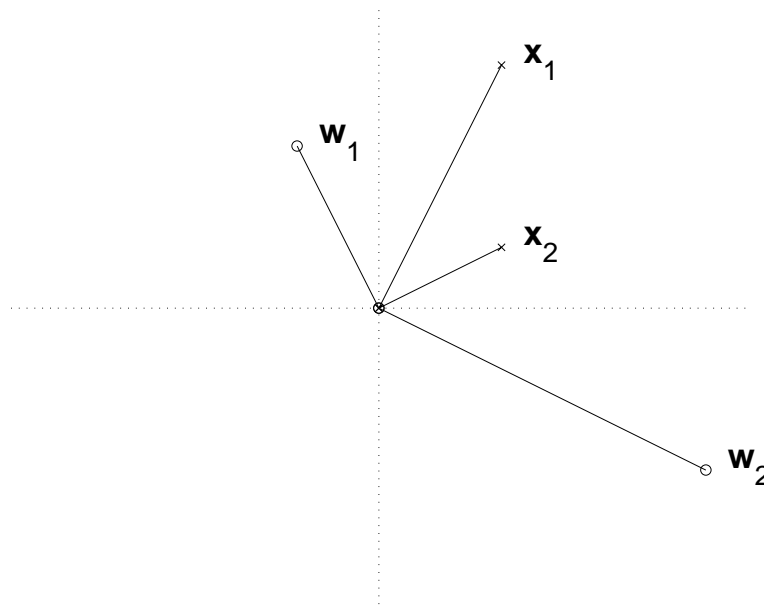
The output, therefore, is merely the projection of a given vector, \mathbf{x} , onto the weight vector \mathbf{w} . Learning then becomes the search for a particular direction in the data space which satisfy the criteria of the BCM equations. This way of thinking about the modeling will become increasingly useful in this work.

1.5.1 Fixed Points

The fixed point will be found when $E[\dot{\mathbf{w}}] = 0$, or in other words, $y = 0$ or $y = \theta$. That gives us four possibilities.

- $(\mathbf{w} \cdot \mathbf{x}_1) = 0$ and $(\mathbf{w} \cdot \mathbf{x}_2) = 0$: zero weight vector
- $(\mathbf{w} \cdot \mathbf{x}_1) = 0$ and $(\mathbf{w} \cdot \mathbf{x}_2) = \theta \neq 0$: weight orthogonal to \mathbf{x}_1 and has projection θ on \mathbf{x}_2
- $(\mathbf{w} \cdot \mathbf{x}_1) = \theta \neq 0$ and $(\mathbf{w} \cdot \mathbf{x}_2) = 0$: weight orthogonal to \mathbf{x}_2 and has projection θ on \mathbf{x}_1
- $(\mathbf{w} \cdot \mathbf{x}_1) = \theta \neq 0$ and $(\mathbf{w} \cdot \mathbf{x}_2) = \theta \neq 0$: weight has equal projections, θ , on both \mathbf{x}_1 and \mathbf{x}_2

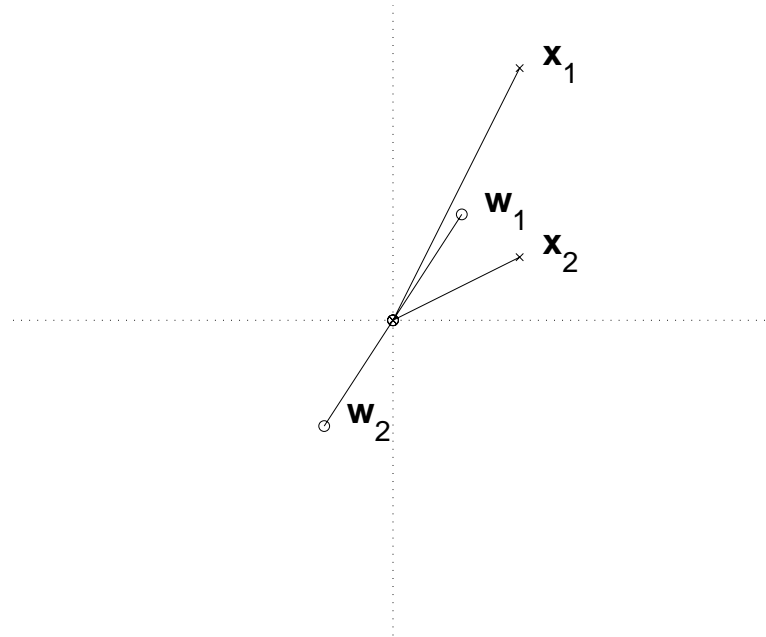
It can be shown (Section A.5.4) that the only two *stable* fixed points are the ones where the weight vector is orthogonal to all but one of the inputs. For the input patterns in the previous picture the stable solutions are shown here



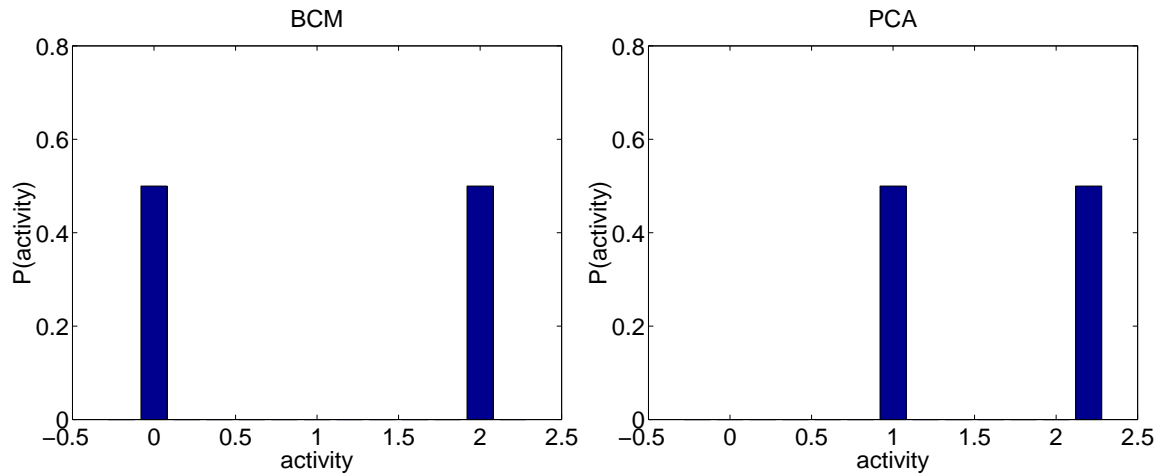
The solution \mathbf{w}_1 is *selective* to input pattern \mathbf{x}_1 , because it responds to (has a non-zero projection on) \mathbf{x}_1 , and does not respond to (has a zero projection on) \mathbf{x}_2 . Similarly, \mathbf{w}_2 is *selective* to input pattern \mathbf{x}_2 .

1.6 Two Dimensional Model: PCA

The simple two dimensional case also highlights some of the properties of the PCA learning rule. Given two input patterns, \mathbf{x}_1 and \mathbf{x}_2 as before, it is a simple calculation to get the fixed points (see Section A.4), when we remember that the neuron is seeking directions of *maximum variance*. An example is shown here



It is clear that the BCM neuron and the PCA neuron are converging to very different fixed points, but the primary difference between the two may not be completely clear. To help elucidate this difference further, it is helpful to look at the output distribution of the neuron at the fixed point for each of these learning rules. The output distribution is merely the probability density for finding a particular output of the neuron over the entire input environment. This function tells us much about the input environment, and about the properties of the learning rule.



From this picture, it appears (at least roughly) that the PCA rule is trying to have most of its responses strong, while the BCM rule is trying to have a small subset of its responses strong, and the others weak. We will make this more precise in Chapter 3.

1.7 LTP and LTD

The experimental verification of synaptic plasticity has occurred primarily in the hippocampus, one of the oldest areas of the brain. In this area the circuitry is relatively simple, and well understood, which makes experimental manipulation easier. Some of the work has more recently been reproduced in visual cortex, as well as many other cortical areas.

Some twenty years after Hebb proposed a theoretical process for storing memories (Hebb, 1949), such a process was found experimentally (Bliss and Lømo, 1973; Bliss and Gardner-Medwin, 1973). The experiment consisted of taking a pathway in the brain and stimulating it with high frequency electrical pulses (Figure 1.6, center). The pathway was tested, both before and after the high frequency stimulation, with very low frequency “baseline” stimulation. It was shown that the response to baseline stimulation after the high frequency input was both much *higher* than before, and very *long lasting* (on the order of weeks in awake animals). Such an increase, referred to as long term potentiation (LTP), in our interpretation is a result of increased synaptic efficacy.

Guided by the BCM theory, similar experiments were performed, but used *low frequency* input instead of high frequency. As was predicted by the BCM rule, a *decrease* in response followed the stimulation (Figure 1.6, bottom). The decrease was long lasting, and was therefore referred to as long term depression (LTD). An experiment was then formulated to measure the BCM $\phi(\cdot)$ function directly.

If the cell activity is some integrated activity, over a small time window, then it would be proportional to the input frequency. If also, in the short time of the input stimulus we assume that the threshold doesn’t move much, then the $\phi(\cdot)$ function would be approximately constant. The total weight

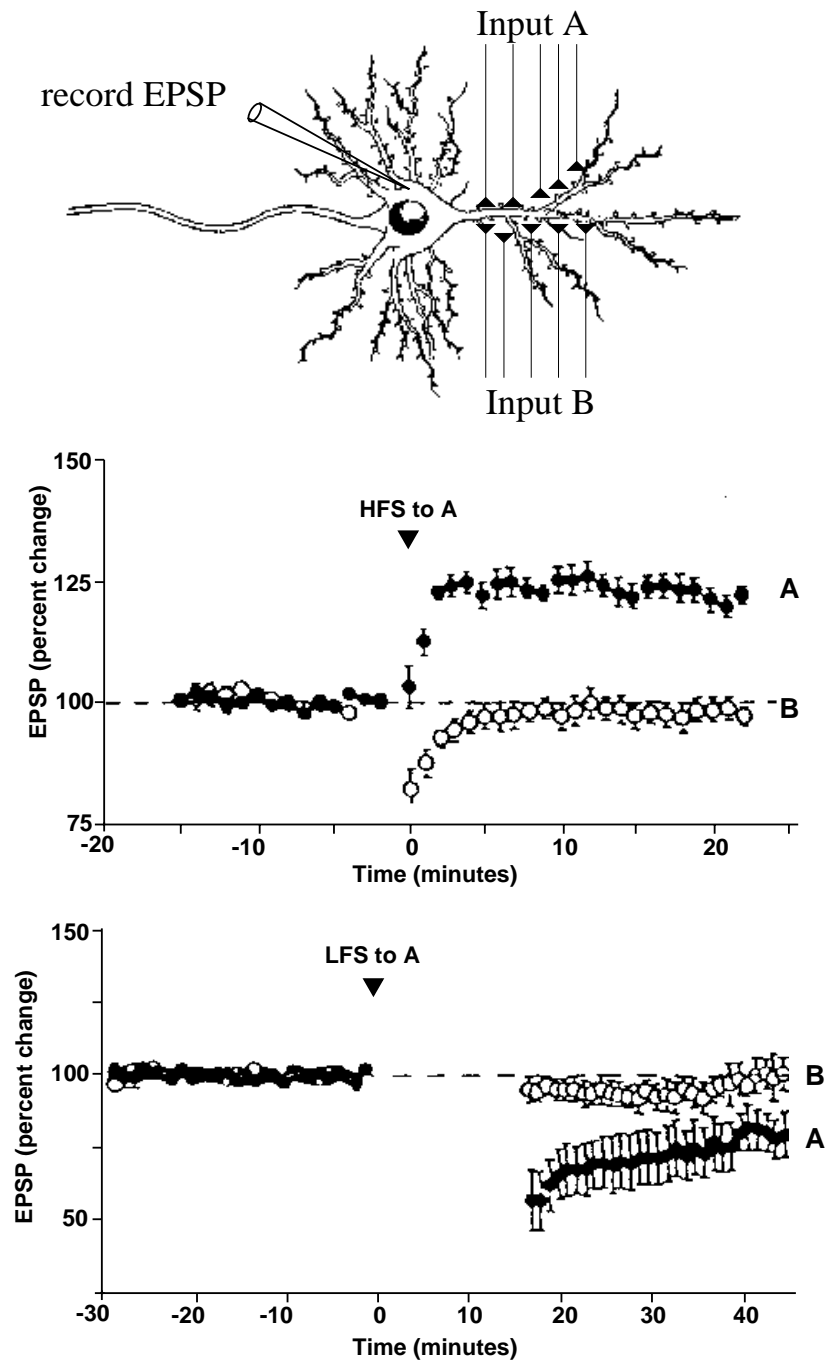


Figure 1.6: LTP and LTD (results from Kirkwood and Bear (1995)). Very low frequency “baseline” stimulus is presented alternately to two independent pathways, A and B. Measurements of excitatory postsynaptic potentials (EPSPs, or simply, the activity of the cell) are performed. After high frequency stimulation (HFS) to pathway A, the response of that pathway is enhanced (LTP) and independent pathways, B, are unaffected. Low frequency stimulation (LFS) to pathway A causes a *reduced* response (LTD) in that pathway, leaving independent pathways unaffected. These forms of LTP and LTD are often called *homosynaptic* LTP/LTD, referring to the fact that only the stimulated pathway is affected.

change for N input spikes at a particular frequency would be

$$\Delta w = \sum_{i=1}^N \phi(y, \theta) x_i \quad (1.27)$$

$$= \phi(y, \theta) \sum_{i=1}^N x_i \quad (1.28)$$

$$= \phi(y, \theta) \times (\text{total input}) \quad (1.29)$$

Since the total input into the cell is kept constant (N spikes for all input frequencies), the measured change in synaptic efficacy, or equivalently, the amount of LTP/LTD is proportional to $\phi(y, \theta)$. This is strictly true only if the updates on the weights occur *after* the presentation of the entire input set. If the weights modify during the input set, then the value of the output, y , would change too and we couldn't pull the $\phi(\cdot)$ function out of the sum. A quick simulation shows that if the weights modify instantaneously, but have a reasonably small learning rate, the same result occurs.

Figure 1.7, top, shows the result of the experiment. The resulting curve looks remarkably like the BCM modification function. In order to test the presence of a moving threshold, the same measurement was performed, this time a group of rats with no visual experience was compared to a group with normal visual experience (Figure 1.7, bottom). The observed shift was consistent with the predicted motion of the modification threshold, θ .

1.7.1 Discussion

Although the experimental evidence for the BCM theory is compelling, we must keep in mind that it is not the only interpretation of the data. The “sliding threshold” result, for example, could be caused merely by a general increase in excitability caused by normal vision, and not part of the learning rule. Since there is no frequency (from Figure 1.7) at which LTD changes to LTP with visual experience, a simple scaling could be the cause of the apparent shift. Unfortunately, those data points which lie near enough to the axis to shift from LTD to LTP are those very points which are extremely difficult to measure accurately. It is true, though, that the basic assumptions about the interpretation of the variables \mathbf{x} and y , and the proposal of a particular form of synaptic modification, has led to experiments which are consistent with BCM.

One clear difference between the BCM rule and the PCA rule is the result when there is *no input* into a particular pathway. From the PCA rule, we would predict that LTP induced on one pathway would cause LTD of silent pathways, because the output of the cell would be positive and thus the decay term would be the remaining term for any silent synapse. LTD caused by the stimulation of an independent pathway is called *heterosynaptic LTD*, and has been measured in hippocampus (Christie and Abraham, 1992). This is compared to the LTD presented earlier, which is commonly called *homosynaptic LTD*. Currently there is little known about the connection of *either* of these two forms of LTD to behavior or changes in selectivity. A nice discussion of these topics can be found in Dudek (1996).

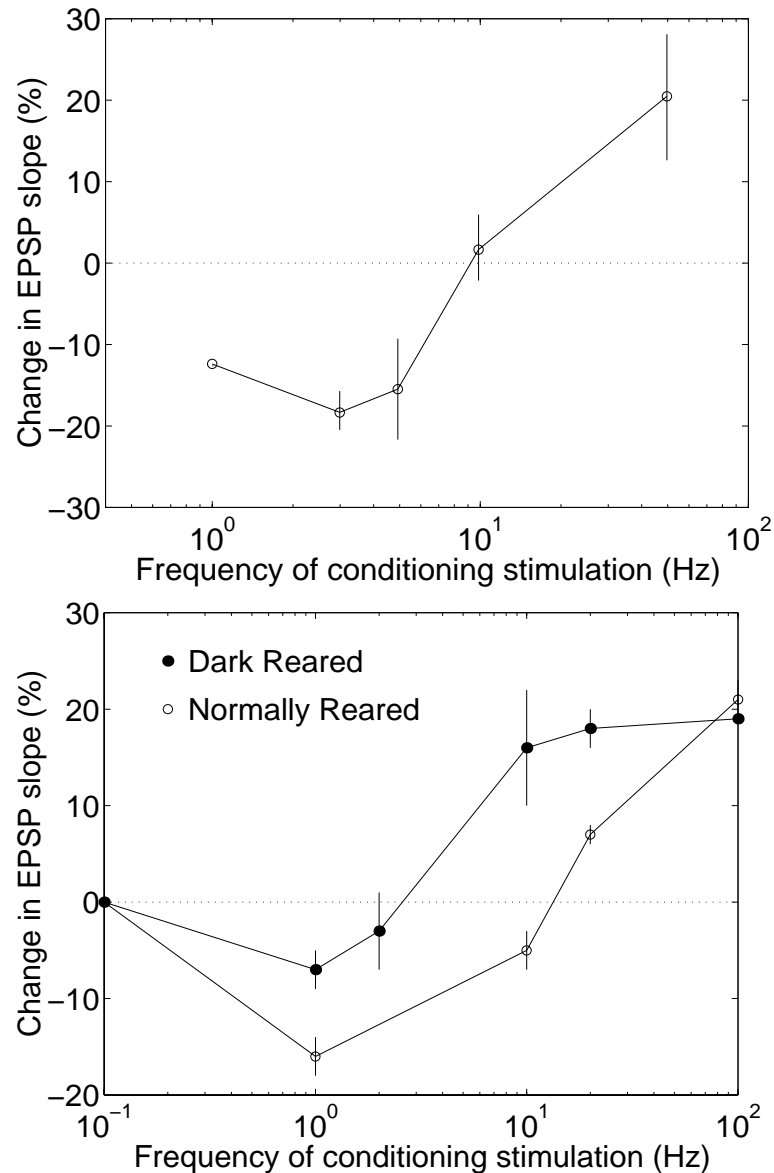


Figure 1.7: Measuring the BCM $\phi(\cdot)$ function. Shown is the change in EPSP (excitatory postsynaptic potential, e.g. activity) as a function of the input frequency. Shown above is from Dudek and Bear (1992), measured in hippocampus. The total input into the cell is kept constant, for all input frequencies, so the change in activity (or synaptic efficacy) is a direct measure of the modification function, $\phi(\cdot)$. Shown below is Kirkwood et. al. (1996). The two graphs correspond to the same measurement performed on rats with no visual experience (dark reared) and those with normal visual experience (normally reared). An activity dependent shift is observed, which is consistent with the motion of the modification threshold, θ .

Another way to differentiate between these rules is to look at the dynamics of synaptic modification in different conditions. One experiment used a culture of neurons, applied various pharmacological agents to modify the activity of the neurons and measured changes in the synaptic efficacy (Turrigiano et al., 1998). They found that reducing the activity (with TTX) for 1-2 days caused an overall *increase* in the synaptic efficacy, while increasing the activity (with bicuculline) for 1-2 days caused a *decrease* on the synaptic efficacy. This modulation seemed to be performed to bring the activity level back to a particular value. This type of scaling is consistent with the weight scaling of BCM, as the threshold moves to maintain the cell activity. If the observed synaptic changes are a result of the BCM mechanism, then it implies the threshold moves at most on the order of *hours*. This is consistent with the calculation of this time scale performed in the next chapter, using a completely different approach.

An experiment using priming in rat hippocampus (Holland and Wagner, 1998) observed changes consistent with the BCM rule, and the motion of the threshold on the order of an hour. In this experiment they gave a stimulus, called a priming stimulus, which made subsequent LTD protocols more effective. Then they measured how long after the priming LTD was more effective, and by how much. In our interpretation, the priming stimulus moved the threshold up to a high values, and then the threshold moved back down over time. This would correspond to the time over which LTD would be more effective, which the experimenters measure to be on the order of an hour.

Another way of differentiating between learning rules, using the dynamics, is to make use of the oscillations. The oscillation of the sliding threshold is unique to the BCM learning rule, and thus could be a strong support of the theory if measured. Yet another method is to use the rules to predict the results of visual deprivation experiments, which is the topic of the next chapter.