Chapter 4

Correlational Models

4.1 Introduction

Two of the most influential recent models of the development of the visual system are those of [Linsker 1986] (see also [Linsker 1988]) and [Miller et al 1989]. Linsker considered the maturation of a number of layers of units with weights between successive layers, modelling very roughly the layers in the visual pathway from one eye to V1. Miller considered a two-layer binocular system of two presynaptic sheets (which he identified with the lateral geniculate nuclei) innervating a single postsynaptic sheet (the cortex). Linsker's primary concern was the development of orientation selectivity given both excitatory and inhibitory weights, whereas Miller's was the development of ocular dominance given two purely excitatory projections. However, though developed independently, there is a close relationship between these models. In fact, Miller's model (in its simplest form) is the natural way of expressing Linsker's model in the binocular case. Each is based on a type of Hebbian learning that Miller refers to as a *correlational* rule: the form of receptive field development is determined by the correlations between presynaptic units. These correlational models have generated much interest, partly because the form of the learning rule makes them susceptible to extensive mathematical analysis. It is thus important for us to attempt to make clear the behaviour of these models (particularly Miller's binocular model) with regard to their position along the dimensions of interest identified in the introduction.

Miller argues that the important parameters of his model are both biologically meaningful and experimentally measurable in the natural system. His model produces good stripe morphology, and can account for the effects of monocular deprivation (although we will return to this point later). Thus he states "In summary, a simple correlation-based plasticity rule, combined with a very simple model of geniculate correlations, geniculocortical connectivity, and cortical connectivity or interactions, is sufficient to account for a large part of the observed phenomenology of ocular dominance segregation" [Miller 1990(a)].

Miller and Linsker both assume a fixed topography in their models. We leave further discussion of whether they could produce topography to the Conclusions chapter. The purpose of this chapter is to investigate to what extent the form of correlations within and between eyes influences the development of monocularity in Miller's model. In his analysis and simulations, Miller assumed that activity in the two eyes is described by identical correlation functions. In addition, in his simulations he usually assumed that there was zero or negative correlation between the eyes. We start by reviewing Linsker's model of receptive field development, the recent eigenvector analysis of this model [MacKay & Miller 1990(a), MacKay & Miller 1990(b)], and its relationship to Miller's model for the development of ocular dominance. We then analyse a more general situation: the addition of small correlations that are different in the two eyes, and of small correlations between the two eyes. We do this by means of a perturbation analysis of the general two-eye correlation matrix. A preliminary account of this work is in [Dayan & Goodhill 1992].

4.1.1 The assumptions behind correlational models

The most important assumptions underlying the models of Linsker and Miller are as follows (for a derivation of rules in this form by the simplification of general Hebbian rules, see [Miller 1990(b)]):

- The output of a post-synaptic unit is a linear weighted sum of the activity on each input line multiplied by the weight on that line.
- During learning, each weight is updated in proportion to the product of pre- and post-synaptic activity on that line.
- Learning is sufficiently slow that we can assume changes in the weights are insignificant over the time taken to present a subset of the input patterns large enough to accurately represent their internal statistics.

These three assumptions allow us to express the learning rule in terms of the covariance matrix of the input activities, which can then be analysed in terms of its eigenvectors. These are the weight structures that develop independently in the system, and govern the ultimately stable weight pattern. Since the behaviour of correlational models is to a large extent governed by the eigenvectors of the covariance matrix, they bear a close relationship with Principal Components Analysis. However, it can be the case in correlational models that finite saturation limits lead to subprincipal eigenvectors determining the final weight pattern [MacKay & Miller 1990(a), MacKay & Miller 1990(b)].

It is instructive to compare these assumptions with those of the Neural Activity model [Willshaw & von der Malsburg 1976] as discussed in chapter 3. Although in the Neural Activity model the *input* to a postsynaptic unit is a linear weighted sum of input activities, there is a threshold non-linearity in the response of units. Learning also occurs in the Neural Activity model by a Hebbian mechanism, but weights are updated every time an input pattern is presented: it is not assumed that learning is much slower than the rate of pattern presentation. These differences mean that the Neural Activity model does not succumb to the same type of analysis as the correlational models of Miller and Linsker.

4.2 Mathematics of correlational models

4.2.1 Linsker's model

[Linsker 1986] considered a system of several layers of units that mature sequentially, with feed-forward connections only between adjacent layers. There are no connections between units in the same layer (except for the last layer which we will not consider), so that each unit develops independently. A predefined topography is assumed: each unit in layer $\mathcal{M}$ is connected to units in the previous layer $\mathcal{L}$ with a probability given by a gaussian function of the distance in $\mathcal{L}$ from the topographically matching position. To speed computation, on occasion Linsker used a simpler version of this gaussian spread of randomly located synapses: he assumed a regular grid of 'representative' synapses, the strength of each scaled by a gaussian arbor function. This is equivalent to the first case when there are enough synapses so that random fluctuations are not significant. However, it complicates the analysis slightly, since the covariance matrix Q is now multiplied by the arbor matrix A, and the product QA is not in general symmetric.

Linsker's equation is now derived (following [MacKay & Miller 1990(b)]). Consider a single unit in layer $\mathcal{M}$ which has connections of strength w_i to a group of units in layer $\mathcal{L}$ with activities $x_i^{\mathcal{L}}$. The activity of this unit in $\mathcal{M}$, $x^{\mathcal{M}}$, is calculated as

$$x^{\mathcal{M}} = \sum_{i \in \mathcal{L}} w_i x_i^{\mathcal{L}} + a_1 \tag{4.1}$$

where a_1 is a constant. Linsker assumed that a_1, and all other constants he introduced, are the same for all units within a particular layer. He used the following Hebbian rule for weight-updating (see the second assumption above):

$$\dot{w}_i = (x^{\mathcal{M}} - c_2)(x_i^{\mathcal{L}} - c_1) + c_3 \tag{4.2}$$

where c_1, c_2 and c_3 are constants. We can now substitute for $x^{\mathcal{M}}$ from (4.1) into (4.2):

$$\dot{w}_i = \left(\sum_{j \in \mathcal{L}} w_j x_j^{\mathcal{L}} + a_1 - c_2 \right) (x_i^{\mathcal{L}} - c_1) + c_3$$

From the third assumption (see page 38), we can average over the input patterns $\mathbf{x}^{\mathcal{L}}$ to obtain an equation for the average rate of change of w_i, in terms of the mean activities $\overline{x}_i^{\mathcal{L}}$ in layer $\mathcal{L}$ and the covariance matrix $Q_{ij}^{\mathcal{L}} \equiv \langle (x_i^{\mathcal{L}} - \overline{x}_i^{\mathcal{L}})(x_j^{\mathcal{L}} - \overline{x}_j^{\mathcal{L}}) \rangle$, where angle brackets denote a time average. If we also assume that the first-order statistics of the patterns in layer $\mathcal{L}$ are uniform, i.e. $\overline{x}_i^{\mathcal{L}} = \overline{x}^{\mathcal{L}}\ \forall i$, we can combine constants to obtain the simple equation

$$\dot{w}_i = \sum_{j \in \mathcal{L}} (Q_{ij}^{\mathcal{L}} + k_2) w_j + k_1$$

or in vector form

$$\dot{\mathbf{w}} = (Q + k_2 J)\mathbf{w} + k_1 \mathbf{n} \tag{4.3}$$

where J is the matrix $J_{ij} = 1,\ \forall\, i, j$, and $\mathbf{n}$ is the 'DC' vector $n_i = 1,\ \forall\, i$. Note that this is slightly different from Linsker's original equation, due to a rescaling of the constants k_1 and k_2 to avoid implicit dependence on the number of synapses appearing in (4.3).

The stability of this equation depends on the eigenvalues of the matrix $Q + k_2 J$. The fixed point $\mathbf{w}_{FP}$ of the equation is given by

$$\mathbf{w}_{FP} = -k_1 (Q + k_2 J)^{-1} \mathbf{n},$$

and it can be shown (e.g. [MacKay & Miller 1990(b)]) that each eigenvector of the matrix $Q + k_2 J$ is a weight structure that grows exponentially relative to the fixed point, with rate given by the appropriate eigenvalue. Since $Q + k_2 J$ is symmetric, it has a complete orthonormal set of eigenvectors with real eigenvalues (see e.g. [Horn & Johnson 1985]). The fixed point is unstable if there exist eigenvectors with positive eigenvalues, and in this case the eigenvector with the largest eigenvalue will dominate as $t \to \infty$. In order to keep the weights within bounds, Linsker used saturation: once weights reach upper or lower limits they are 'frozen' and no further change is allowed. He used two forms of saturation rule. The first assumed that there exist two statistically indivisible populations of units in layer $\mathcal{L}$, one excitatory and one inhibitory. The saturation limits for each unit were then taken to be 0 and 1. This was later replaced with a single type of unit, and weights that could take on both positive and negative values. This is formally equivalent, though somewhat less biologically plausible, given Dale's law that cells in the natural system are either excitatory *or* inhibitory. Usually he considered weights in the range $[-0.5, 0.5]$. It can be shown [Linsker 1986] that all, or all but one, of the weights saturate at the upper or lower limits.

The regime of activity considered was uncorrelated random firing in the first layer $\mathcal{A}$. This leads to gaussian correlations in layer $\mathcal{B}$ (due to the assumption of a gaussian synaptic density function) and for each subsequent layer the correlations can be calculated for that layer once the weights have matured.

Linsker reported (for carefully chosen parameter values that varied from layer to layer) all-excitatory or all-inhibitory receptive fields in layer $\mathcal{B}$ (the second layer), centre-surround units in layers $\mathcal{C}$ to $\mathcal{F}$, and eventually oriented units in layer $\mathcal{G}$. Oriented units can in fact be obtained in layer $\mathcal{C}$ for appropriate parameter values, but these units are less robust than those that appear in layer $\mathcal{G}$.

4.2.2 The eigenvector analysis

In a Hebbian rule such as (4.3), without any constraints, the eigenvector with the largest positive eigenvalue will eventually dominate. It is thus clear that an eigendecomposition of the covariance matrix can reveal much about the development of the system. However, matters are complicated by the introduction of saturation constraints. Firstly, non-linear phenomena may now occur once the weights start to reach the saturation limits, and secondly the initial conditions may give some eigenvectors a 'head start' such that, despite not growing as fast as other eigenvectors, they are still the first to reach the saturation limits. [MacKay & Miller 1990(a), MacKay & Miller 1990(b)] performed the eigendecomposition of Linsker's covariance matrix $Q + k_2 J$, and analysed under what circumstances the principal eigenvector determined the final weight structure. This is somewhat similar to the analysis of the Tea Trade model performed by [Häussler & von der Malsburg 1983], except that Häussler and von der Malsburg also consider the effect of higher-order terms, which Mackay and Miller do not (see also [Turing 1952]). Mackay and Miller made the following important assumption:

> The principal features of the dynamics are established before the hard limits are reached. When the hypercube is reached, it captures and preserves the existing weight structure with little subsequent change. [MacKay & Miller 1990(b)].

Thus they ignored the possibility of non-linear effects after weight saturation.

They used the standard wave-vector notation of Quantum Mechanics (see e.g. [Schiff 1968]) to describe the eigenvectors. Under most circumstances, they found that the following receptive field structures are the principal eigenvectors:

- 1s: An all-excitatory or all-inhibitory receptive field, which has the largest possible DC component of all the eigenvectors.
- 2s: A centre-surround receptive field: a core of positive (negative) components surrounded by a ring of negative (positive) components. The sum of the components is not generally equal to zero, and thus this structure has a medium-sized DC component.
- 2p: An oriented receptive field: a region of positive components separated from a region of negative components by a straight line. The two regions are equally balanced, and thus this structure has zero DC component.

In this chapter we label 2p structures as x_i, and for the moment do not distinguish between the eigenvectors with non-zero DC, labelling them as y_i. Mackay and Miller found that the three varieties of eigenvector listed above are usually the three principal components of Linsker's Q matrix. However, whether the eigenvector with the highest eigenvalue wins depends on the parameters k_1 and k_2:

- k_1 does not affect the eigenvalues, but biases the competition between zero DC and non-zero DC eigenvectors. A positive (negative) value of k_1 gives a head start to eigenvectors with positive (negative) DC component.
- k_2 determines which is the principal eigenvector. Modes x with zero DC (i.e. $x.n = 0$) have eigenvalues that are constant with k_2, while modes y with non-zero DC (i.e. $y.n \neq 0$) have eigenvalues that increase monotonically with k_2.

The effect of this for various combinations of k_1 and k_2 is shown in [MacKay & Miller 1990(a), MacKay & Miller 1990(b)].

4.3 Extension to two eyes

We now consider the natural extension of Linsker's equations to the case of two eyes, or more generally two populations (for instance on-off and off-on centre-surround cells: for concreteness we will express results in terms of two eyes). This leads to a system more general than that studied by Miller: it is shown that Miller's model can be regarded as a special case of this. We consider a two-layer system, which in general will correspond to Linsker's $\mathcal{B}$ to $\mathcal{C}$ connections, with gaussian correlations in layer $\mathcal{B}$ (the "input" layer).

A correlational rule for two populations can now be found in the same way as the monocular derivation given above:

$$\begin{pmatrix} \dot{w}_L \\ \dot{w}_R \end{pmatrix} = \begin{pmatrix} Q_L + k_2 J & Q_C + k_2 J \\ Q_C + k_2 J & Q_R + k_2 J \end{pmatrix} \begin{pmatrix} w_L \\ w_R \end{pmatrix} + k_1 \begin{pmatrix} n \\ n \end{pmatrix} \tag{4.4}$$

where Q_L gives the covariance between units within the left eye, Q_R gives that between units within the right eye, and Q_C the covariance between units in the two eyes. We assume Q_C is symmetric: this is equivalent to saying that the left and right eyes see the same set of patterns. Define Q_* as the full two eye iteration matrix. We constrain all weights to be positive, for reasons that will be apparent later.

4.3.1 Miller's special case

The above equation with $k_1 = k_2 = 0$ is the basis of Miller's model, as we now show. His equation for the development of the left-eye weights is:

$$\begin{aligned} \dot{w}^L(x, \alpha, t) &= \lambda A(x-\alpha) \sum_{y,\beta} I(x-y)[C^{LL}(\alpha-\beta) w^L(y, \beta, t) + C^{LR}(\alpha-\beta) w^R(y, \beta, t)] \\ &\quad - \gamma w^L(x, \alpha, t) - \epsilon^L A(x-\alpha) \end{aligned}$$

where α, β index over retinal units and x, y index over cortical units, A is the arbor function, I is related to the cortical interaction function, C^{LL} and C^{LR} are the correlation functions within the left eye and between the left and right eyes respectively, and the last two terms are decay terms (which Miller usually took to be zero). There is an analogous equation for the right eye weights.

If we consider just the development of monocularity for an individual cortical unit, we can ignore the cortical interaction function. Ignoring also the arbor function (which in practice appears to have little effect) and setting the decay terms to zero yields

$$\begin{aligned} \dot{w}^L(\alpha, t) &= \lambda \sum_{\beta} [C^{LL}(\alpha-\beta) w^L(\beta, t) + C^{LR}(\alpha-\beta) w^R(\beta, t)] \\ \dot{w}^R(\alpha, t) &= \lambda \sum_{\beta} [C^{RR}(\alpha-\beta) w^R(\beta, t) + C^{RL}(\alpha-\beta) w^L(\beta, t)] \end{aligned}$$

Or, in matrix notation (taking $\lambda = 1$),

$$\begin{aligned} \dot{w}^L &= C^{LL} w^L + C^{LR} w^R \\ \dot{w}^R &= C^{RR} w^R + C^{RL} w^L \end{aligned}$$

E-vector	E-value	Conditions	
(x_i, x_i)	λ_i	$Qx_i = \lambda_i x_i$	$x_i.n = 0$
$(x_i, -x_i)$	λ_i	$Qx_i = \lambda_i x_i$	$x_i.n = 0$
$(y_i, -y_i)$	μ_i	$Qy_i = \mu_i y_i$	$y_i.n \neq 0$
(z_i, z_i)	ν_i	$(Q + 2k_2 J)z_i = \nu_i z_i$	$z_i.n \neq 0$

Table 4.1: The eigendecomposition of Q_*^m.

These are the same equations as expressed in (4.4), with k_1 and k_2 equal to zero. The absence of k_2 is equivalent to assuming that the average activity of each input unit is zero [Miller 1990(a)].

These equations are coupled. However, Miller assumed that $C^{LL} = C^{RR}$ and $C^{LR} = C^{RL}$ (similarly to [Swindale 1980]), and was thus able to separate the equations into sum ($S^S = w^L + w^R$) and difference ($S^D = w^L - w^R$) variables, which evolve independently. S^D describes the development of ocular dominance, and Miller showed that with saturation, bottom-limited at zero, and a subtractive normalization rule to limit synaptic strengths (an issue to which we return later), the patterns of dominance between the two eyes are determined by the initial weight values and the fastest-growing components of S^D. The starting condition Miller adopts has $w^L - w^R = \epsilon' p$ where ϵ' is small, and $w^L + w^R = q$, which is $\mathcal{O}(1)$.

4.3.2 The eigendecomposition of Miller's matrix

In his simulations, Miller usually took $Q_C = 0$, i.e. the case of no correlations between the two eyes. Using this and the assumption that $Q_R = Q_L = Q$, we can calculate the eigendecomposition of the resulting matrix Q_*^m in the same way that Mackay and Miller analysed the eigendecomposition of the one-eye matrix $Q + k_2 J$ (see table 4.1).[1] There are two basic types of eigenvectors: $e_1 = (u, u)$ and $e_2 = (u, -u)$. These can be further subdivided depending on the form of u. If u has zero DC, then both e_1 and e_2 correspond to "striped" binocular patterns (oscillation between eyes occurring within a receptive field): we refer to these as x. In this case e_1 and e_2 are degenerate, and thus any linear combination of them is also an eigenvector of Q_*^m. We see below how the introduction of perturbative correlations can break this degeneracy. An eigenvector u with non-zero DC (y and z) is also binocular if u still has both positive and negative components (e.g. a 2s centre-surround pattern). The only eigenvector that corresponds to a monocular pattern is e_2 when u is entirely DC (i.e. 1s). Note that since weights are positive in Miller's system, the dominance of e_2 for u entirely DC leads to a final weight pattern of $(u, 0)$, i.e. monocular. Monocularity could not develop in the context of Linsker's saturation limits even if eigenvectors of the form $(u, 0)$ existed, because growth of sub-principal eigenvectors always ensures that all weights eventually saturate at their limiting values. Thus referring to table 4.1 we see that the only monocular solution is $(y_i, -y_i)$ for i such that y_i is 1s. [Miller 1990(a), pages 296–300] compares empirically the development of various eigenvectors, including circumstances under which the monocular eigenvector does and does not dominate.

4.4 Perturbation analysis

We now return to the complete two eye equation (4.4), and analyse a more general case than that considered by Miller. In particular we do not assume $Q_L = Q_R$: Miller's device of decomposing equation (4.4) into sum and difference variables is no longer valid, and more general methods are required. We also do not assume that $k_1 = k_2 = 0$. We perform the eigendecomposition of the general matrix Q_* with the assumption that $Q_L - Q_R$ and Q_C are small. This amounts to a perturbation analysis of Miller's special case, and serves two functions:

- It provides a step in the direction of understanding the general two eye system.

[1] We do not describe the case of degeneracy between modes of different index i. This is for notational convenience, and does not affect the results. Also for convenience we write (x_i, x_i) for $(x_i, x_i)^T$, *etc.*

- It explains some of the empirical results found by Miller, reveals problems of stability in his model, and predicts the effects of positive correlations between eyes.

Let $Q_L = Q + \epsilon Q_1$, $Q_R = Q + \epsilon Q_2$ and $Q_C = \epsilon Q_c$, and call the resulting matrix Q_*^ϵ:

$$Q_*^\epsilon = \begin{pmatrix} Q + \epsilon Q_1 + k_2 J & \epsilon Q_c + k_2 J \\ \epsilon Q_c + k_2 J & Q + \epsilon Q_2 + k_2 J \end{pmatrix} \tag{4.5}$$

It is now of interest to calculate the effects of these perturbations on the eigenvectors and eigenvalues, particularly with regard to the following points:

- Different eigenvectors may be perturbed by different amounts, possibly changing the nature of the competition between them.
- Some eigenvectors may not be stable to the perturbation, *i.e.* if (u_1, u_2) is an eigenvector of Q_* it may not be the case that $(u_1 + \epsilon a_1, u_2 + \epsilon a_2)$ is an eigenvector of Q_*^ϵ for some vectors a_1 and a_2.
- Previously degenerate eigenvectors may split, implying one is more favoured in the race to saturation. This phenomenon is well known in Quantum Mechanics (see e.g. [Schiff 1968]).

To calculate the eigenvalues and eigenvectors of the perturbed system, we consider the following equation:

$$\begin{pmatrix} Q + \epsilon Q_1 + k_2 J & \epsilon Q_c + k_2 J \\ \epsilon Q_c + k_2 J & Q + \epsilon Q_2 + k_2 J \end{pmatrix} \begin{pmatrix} \alpha_i u_i + \epsilon a_i \\ \beta_i u_i + \epsilon b_i \end{pmatrix} = (\lambda + \epsilon \mu_i) \begin{pmatrix} \alpha_i u_i + \epsilon a_i \\ \beta_i u_i + \epsilon b_i \end{pmatrix} \tag{4.6}$$

where u_i is a general eigenvector of Q with eigenvector λ. That is, we look for solutions of the form $(\alpha_i u_i + \epsilon a_i, \beta_i u_i + \epsilon b_i)$. Consideration of zero-order terms gives the results in table 4.1: thus u_i is an eigenvector of Q. Calculating terms of $\mathcal{O}(\epsilon)$ gives the following two equations (where for clarity we drop the subscript i):

$$(Q - \lambda I)a + k_2 J(a + b) = -(\alpha Q_1 + \beta Q_c - \alpha\mu I)u \tag{4.7}$$

$$(Q - \lambda I)b + k_2 J(a + b) = -(\beta Q_2 + \alpha Q_c - \beta\mu I)u \tag{4.8}$$

We now use the fact that since u is an eigenvector of Q, $(Q - \lambda I)p$ can have no components in the direction of u for any p.(This becomes clear if one imagines p expressed in the eigenvector basis: $p = \sum_i a_i u_i$). There are two cases to consider: zero DC and non-zero DC.

Zero DC: $u.n = 0$

For vectors x such that $x.n = 0$, we have $x^T J = 0$ and so multiplying equations (4.7) and (4.8) by x^T makes both left-hand sides vanish, enabling us to write down two expressions for μ. Combining these gives an equation for $\rho = \beta/\alpha$:

$$x^T Q_1 x + \rho x^T Q_c x = x^T Q_2 x + \frac{1}{\rho} x^T Q_c x \tag{4.9}$$

For $Q_c \neq 0$ we define

$$\gamma = \frac{x^T(Q_1 - Q_2)x}{2x^T Q_c x}$$

and thus obtain from (4.9)

$$\rho = -\gamma \pm \sqrt{1 + \gamma^2}$$

Therefore in the perturbed system with $Q_c \neq 0$, the degeneracy is broken: only particular combinations of zero-DC vectors are allowed, namely $(\alpha x, \beta x)$ and $(-\beta x, \alpha x)$ where the ratio of β/α is as given above. The new eigenvalues are

$$\lambda + \frac{\epsilon}{2}\left\{ x^T(Q_1 + Q_2)x \pm \sqrt{[x^T(Q_1 - Q_2)x]^2 + 4[x^T Q_c x]^2} \right\}$$

To explain the meaning of this equation we will return to consider some special cases shortly. For $Q_c = 0$ we refer back to equations (4.7) and (4.8) using $\alpha = 0$, $\beta = 1$, and $\alpha = 1$, $\beta = 0$. This yields the same expression for the perturbation as above, except with $Q_c = 0$.

E-vector	E-value of Q_*^M	Eigenvalue of Q_*^ϵ
$(\alpha_i x_i, \beta_i x_i)$	λ_i	$\lambda_i + \frac{\epsilon}{2}\left\{x_i^T(Q_1+Q_2)x_i + \sqrt{[x_i^T(Q_1-Q_2)x_i]^2 + 4[x_i^T Q_c x_i]^2}\right\}$
$(-\beta_i x_i, \alpha_i x_i)$	λ_i	$\lambda_i + \frac{\epsilon}{2}\left\{x^T(Q_1+Q_2)x_i - \sqrt{[x_i^T(Q_1-Q_2)x_i]^2 + 4[x_i^T Q_c x_i]^2}\right\}$
$(y_i, -y_i)$	μ_i	$\mu_i + \frac{\epsilon}{2} y_i^T(Q_1+Q_2-2Q_c)y_i$
(z_i, z_i)	ν_i	$\nu_i + \frac{\epsilon}{2} z_i^T(Q_1+Q_2+2Q_c)z_i$

Table 4.2: Effect of perturbations on eigendecomposition of Q_*^ϵ.

E-vector	E-value of Q_*^M	Eigenvalue of $Q_*^\epsilon(0,0,Q_c)$
(x_i, x_i)	λ_i	$\lambda_i + \epsilon x_i^T Q_c x_i$
$(x_i, -x_i)$	λ_i	$\lambda_i - \epsilon x_i^T Q_c x_i$
$(y_i, -y_i)$	μ_i	$\mu_i - \epsilon y_i^T Q_c y_i$
(z_i, z_i)	ν_i	$\nu_i + \epsilon z_i^T Q_c z_i$

Table 4.3: Case of $Q_1 = Q_2 = 0$: perturbations only between the eyes.

Non-zero DC: $u.n \neq 0$

For non-zero DC vectors, the $k_2 J(a+b)$ terms in (4.7) and (4.8) do not disappear. However, we can find a value for μ for vectors $(y, -y)$ in table 1 by subtracting (4.7) from (4.8) and substituting $\alpha = -\beta$, and likewise for (z, z) by adding (4.7) and (4.8) and substituting $\alpha = \beta$. The results are summarized in table 4.2, where for clarity we have assumed that x_i, y_i and z_i are normalised (and the subscript i has been replaced).

4.5 Special cases

In order to highlight the effects of these perturbations, we now consider some special cases. This also allows contact to be made with some of the results found empirically by Miller.[2]

4.5.1 Cross-eye perturbation only: $Q_1 = Q_2 = 0, Q_c \neq 0$

We will refer to this case of only perturbative correlations between the two eyes as $Q_*^\epsilon(0,0,Q_c)$. Results are shown in table 4.3. Firstly, we see that all eigenvectors are stable to this perturbation. Secondly, it should be noted that $x^T Q_c x$ is less for zero DC vectors than non-zero DC vectors if we consider all components of Q_c to be of the same sign. We see from table 4.3 that there is a hierarchy of effects on different eigenvectors. For $Q_c < 0$, the eigenvalue of the monocular $(y_i, -y_i)$ for $y_i = 1s$ is increased the most, followed by a smaller increase in the binocular $(x_i, -x_i)$ and $(y_i, -y_i)$ for $y_i \neq 1s$. (x_i, x_i) is decreased slightly, with a bigger decrease in (z_i, z_i). These predictions can be compared with the empirical results of [Miller 1990(a), pages 296–300]. Here, the effect of small-magnitude long-range gaussian anticorrelations between the two eyes on the eigenvalues of (1s,-1s), and (2p,2p) and (2p,-2p), is examined, and it is shown that an increase in eigenvalue for $(y_i, -y_i)$ occurs, while the eigenvalues for $(x_i, x_i), (x_i, -x_i)$ are changed less, as we predict.

[2]There is a technical point that we have not so far discussed. The point of our analysis is to reveal under what circumstances the eigenvalues for different eigenvectors cross. However, it is well known in Quantum Mechanics that perturbation theory breaks down when eigenvalues cross, since the second order terms in the perturbation expansion have differences between the eigenvalues on the bottom, and thus blow up when these become equal (see e.g. [Schiff 1968]). It should be emphasised that this does *not* mean that the eigenvalues do not cross. For the Gaussian correlation and perturbation functions, the first order analysis is correct, and simulation results show that the predictions about crossing are also emprically accurate for random perturbations.

E-vector	E-value of Q_*^M	Eigenvalue of $Q_*^\epsilon(0, 0, Q_c)$
(x_i, x_i)	λ_i	$\lambda_i + \epsilon x_i^T Q_s x_i$
$(x_i, -x_i)$	λ_i	$\lambda_i + \epsilon x_i^T Q_s x_i$
$(y_i, -y_i)$	μ_i	$\mu_i + \epsilon y_i^T Q_s y_i$
(z_i, z_i)	ν_i	$\nu_i + \epsilon z_i^T Q_s z_i$

Table 4.4: Case of $Q_1 = Q_2 = Q_s$, $Q_c = 0$: equal perturbations in each eye.

The effects of these changes depend critically on the relative sizes of the different eigenvectors: it could be that they are so different that perturbations do not affect the competition, or it could be that they are close enough so that perturbations can select a different winning eigenvector from the unperturbed case. We first show the application of this analysis to some numerical results provided by Miller for the case of anticorrelations between the eyes. This also allows us to predict the effects of *positive* correlations between the eyes on Miller's results. Firstly we see from [Miller 1990(a), page 300] that addition of anticorrelations between the eyes leads to an increase in the eigenvalue for the monocular (1s,-1s) solution, and a lesser increase in the eigenvalue for the (2p,2p) and (2p,-2p) solutions, as we predict. Secondly we see that in his third case (relatively short-range correlations within an eye) the addition of positive correlations between the eyes would be enough to cause binocular rather than monocular development of a unit (since reversing the sign of Q_c is equivalent to subtracting the same amount from the eigenvalue of (1s,-1s) that was added before).

We now present simulation results for the case of gradually increasing positive correlations between the two eyes. Parameters are as follows: number of units in LGN = 12×12, within-eye correlations C_w given by e^{-d^2/σ^2} where d is the Euclidean distance between LGN units and $\sigma = 2.0$, between-eye correlations C_b as for within-eye except with $\sigma = 6.0$. Total correlations are given by $C_w + \epsilon C_b$, where ϵ is the perturbation. Figure 4.1 shows the dependence of the eigenvalues on the perturbation for some of the principal eigenvectors. Figure 4.2 shows the principal eigenvector at each magnitude of perturbation. The following effects can be seen.

- The qualitative forms of the eigenvectors are not affected by the perturbation.
- The eigenvalues are affected by the perturbation, in the direction predicted by our analysis.
- At a certain point the eigenvalues of monocular and binocular solutions cross, implying the development of binocular rather than monocular cells.

Thus we see that there are conditions under which perturbations can affect the competition, and that increasing positive correlations between the eyes leads to an increasing tendency to binocularity. These findings suggest an analogy with the results of 'artificial strabismus' experiments [Hubel & Wiesel 1965], where it is found that reducing the correlations between the two eyes leads to increased monocularity. However, the analysis above describes only the case of adding small correlations between two uncorrelated eyes, rather than the case investigated by Hubel and Wiesel of reducing the correlations between strongly correlated eyes.

4.5.2 Within-eye perturbation only: $Q_c = 0, Q_1 = Q_2 = Q_s$

Here all eigenvalues are changed by the same factor, $u^T Q_s u$ (see table 4.4). Note that ρ is in this case a free parameter. However, this can still change the competition, since the eigenvalues of the non-zero DC eigenvectors $(y_i, -y_i)$ and (z_i, z_i) will be perturbed to a greater extent. Miller shows empirically [Miller 1990(a), pages 296–300] that addition of anticorrelations within an eye ($Q_s < 0$) can lead to a 'striped' binocular receptive field. This is as we predict, since the eigenvalue of $(y_i, -y_i)$ is decreased more than those of (x_i, x_i) and $(x_i, -x_i)$.

Some simulation results for the case of gradually increasing long range anticorrelations identical within each eye are shown in figure 4.4. Parameters are as above, except that total correlations are now of the form $C_w + \epsilon C_a$, where C_a represents within-eye anticorrelations with $\sigma = 6.0$. We see that the qualitative forms of the eigenvectors are now continuously distorted by the perturbation. This means that our prediction for

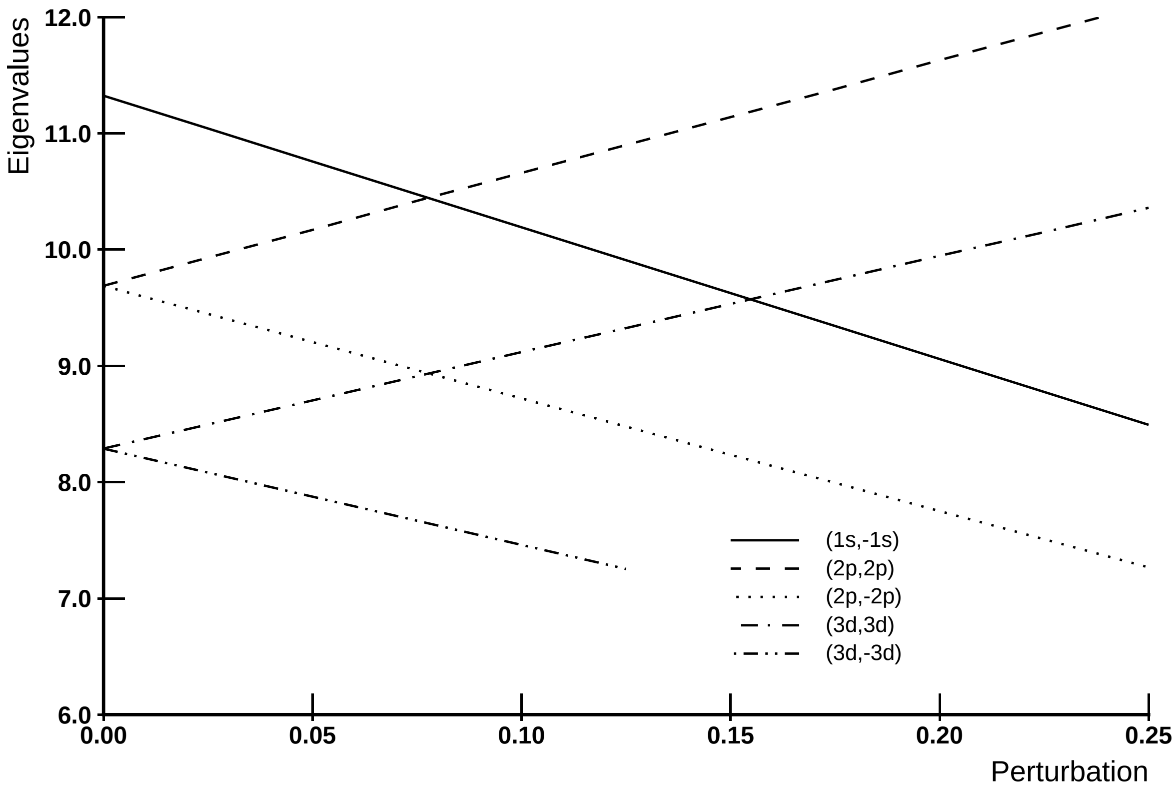

Figure 4.1: Effect of positive cross-eye correlations on the eigenvalues of the principal eigenvectors (not including the (1s,1s) eigenvector, which has largest eigenvalue but is removed by the normalization). Note that the eigenvalues change linearly with the perturbation, and that the eigenvalues cross.