

Stable Orientation Tuning in the Visual Cortex

J.A.F. Heimel^a H. Sompolinsky^b

^a*Dept. of Mathematics, King's College London, The Strand, London WC2R 2LS, UK*

^b*Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem, Israel 91904*

Abstract

We study the behaviour of the Ben-Yishai hypercolumn model [2] under presentation of oriented stimuli, having extended this model by including plastic afferent (LGN to cortex) connections. We find that Hebbian plasticity creates a self-organising map and show that constraining or modifying the standard Hebb rule in a particular way will lead to a contrast-insensitive tuning width, thus giving an explanation for persistent orientation tuning as observed in the visual cortex. Our analytical results confirm those of simulations done by Von der Malsburg [4] and provide a starting point for further analytical treatment of less restricted stimuli.

Key words: orientation tuning, self-organising map, Hebbian learning

1 Introduction

The Hubel and Wiesel (HW) model for explaining orientation-selectivity in the primary visual cortex has been the canonical model for almost forty years. Many features of this model have been confirmed experimentally. However, as early as 1982 Sclar et al. [5] have produced results on contrast-independence of the orientation-selectivity that the HW feed-forward model fails to explain. An elegant model for a single hypercolumn which can account for these data has been studied by Ben-Yishai et al. [2]. It consists of the HW-type afferent connections from the LGN to the cortical hypercolumn in V1, with the added feature of cortico-cortico-interactions of a Mexican hat type, i.e. short-range excitatory and long-range inhibitory. Ben-Yishai et al. showed that the cortical feedback sharpens the tuning and makes the tuning width less sensitive to the contrast of the visual stimulus, in agreement with Sclar's data. The model readily admits a mathematical analysis, making it a perfect starting point

for further research. In the Ben-Yishai model all synapses are fixed. However, in living mammals connections between LGN and the cortex are known to be plastic, even after the development of orientation selectivity. We model this by making the afferent connections subject to Hebbian learning, so that we can study whether Hebbian learning could account for the sustained but flexible set-up of the hypercolumn. Introducing Hebbian learning makes the system a self-organising map, closely related to Amari's neural field model [1] with a different gain function (f-I-relation).

2 The model

The model describes the population rates of a hypercolumn with a circular topology. The activity (firing rate) of populations in the hypercolumn is denoted by $m(\theta)$, where θ , $-\frac{\pi}{2} \leq \theta < \frac{\pi}{2}$ represents the location in the hypercolumn. The evolution of the population rate is ruled by the following mean-field Wilson-Cowan equation:

$$\tau_f \frac{\partial}{\partial t} m(\theta, t) = -m(\theta, t) + \left[\int \frac{d\theta'}{\pi} J(\theta - \theta') m(\theta', t) + I(\theta, t) - T \right]_+, \quad (1)$$

where the linear-threshold function $[x]_+$, which is x for $x > 0$ and zero otherwise, will be our frequency-input relation. The effective threshold of the population is denoted by T and the input conductance of population θ at time t is given by $I(\theta, t) = \mathbf{w}(\theta, t) \cdot \mathbf{x}(t)$. Here \mathbf{x} is the afferent input coming from the LGN and $\mathbf{w}(\theta, t)$ is the synaptic strength vector of population θ . The intra-cortical connection strengths are of the Mexican hat type $J(\theta) = J_0 + J_1 \cos(2\theta)$ and are fixed. The time scale is set by τ_f . In addition to the fast neuronal dynamics, the model features synaptic dynamics in the LGN-cortex connections. This process takes place on a much slower time scale. We take the adiabatic limit and assume that the neurons have reached a stationary state $\bar{m}(\theta, \mathbf{x})$ before a new pattern is presented:

$$\bar{m}(\theta, \mathbf{x}) = \left[\int \frac{d\theta'}{\pi} J(\theta - \theta') \bar{m}(\theta', \mathbf{x}) + I(\theta, \mathbf{x}) - T \right]_+. \quad (2)$$

Learning is considered to have a measurable effect only after the presentation of many input patterns (batch-limit) and is of Hebbian type:

$$\tau_w \frac{\partial}{\partial t} \mathbf{w}(\theta, t) = -\mathbf{w}(\theta, t) + \int d\mathbf{x} p(\mathbf{x}) \mathbf{x} \bar{m}(\theta, \mathbf{x}), \quad (3)$$

where $p(\mathbf{x})$ is the average fraction of time a stimulus \mathbf{x} is shown to the network. Substituting synaptic stationarity into the neuronal stationarity condition (2), we find

$$M(\theta, \mathbf{x}) = \left[\int \frac{d\theta'}{\pi} J(\theta - \theta') M(\theta', \mathbf{x}) + \int d\mathbf{x}' p(\mathbf{x}') \mathbf{x} \cdot \mathbf{x}' M(\theta, \mathbf{x}') - T \right]_+, \quad (4)$$

a fixed point equation for the shape of the stationary activity, given stimulus \mathbf{x} . We will restrict ourselves to patterns \mathbf{x} characterised by an angle ϕ , $-\frac{\pi}{2} \leq \phi < \frac{\pi}{2}$, thus limiting the input coming from the LGN to oriented patterns. To find the solutions of (4) we do not have to specify a particular representation of these patterns, only the inner product between two of them, which for bars or gratings would be something like:

$$\mathbf{x}(\phi) \cdot \mathbf{x}(\phi') = h(\phi - \phi') = h_0 + h_1 \cos(2\phi - 2\phi').$$

3 Results

A hypercolumn, with a one-to-one smooth mapping from angle to population of maximum activity, satisfies $M(\theta, \phi) = M(\theta - \phi)$. This is what we call a perfect map solution and is the configuration postulated in previous work on this model. The specific choice for J and h allows for an easy analysis of the existence of such a solution using only the first couple of its Fourier components. One finds a clipped cosine solution:

$$M(\theta, \phi) = A [\cos 2(\theta - \phi) - \cos 2\theta_c]_+$$

where A , the level of activation, and θ_c , the tuning width, are both functions of $J_0 + h_0$ and $J_1 + h_1$ only. For a stability analysis of the perfect map under neuronal dynamics, the synapses can be taken fixed and one needs to look at the evolution of perturbations of the activity only:

$$\tau_n \frac{\partial}{\partial t} \delta m(\theta, \phi) = -\delta m(\theta, \phi) + \Theta(M(\theta, \phi)) \int \frac{d\theta'}{\pi} J(\theta - \theta') \delta m(\theta', \phi),$$

with the step function $\Theta(x)$ that is 1 if $x > 0$ and 0 otherwise. As outlined in e.g. [3], one easily derives a set of two closed differential equations for the time evolution of the fluctuations, again only using the first couple of Fourier components. We do not study the evolution of fluctuations in the weights of the perfect map directly but via the induced input function. We first calculate

how the stationary activation M changes under the perturbations in the input function I . Using $\delta m = \overline{m}(\theta, \phi) - M(\theta, \phi)$, we find

$$\delta m(\theta, \phi) = \Theta(M(\theta, \phi)) \left\{ \int \frac{d\theta'}{\pi} J(\theta - \theta') \delta m(\theta', \phi) + \delta I(\theta, \phi) \right\}.$$

We then evaluate how the input function reacts to this changed activation:

$$\tau_w \frac{\partial}{\partial t} \delta I(\theta, \phi) = -\delta I(\theta, \phi) + \int \frac{d\phi'}{\pi} h(\phi - \phi') \delta m(\theta, \phi').$$

This coupled system is more difficult to study, but it can be done analytically using restricted Fourier co-efficients

$$\widehat{f}_k(\phi) = \int_{\Phi - \theta_c}^{\Phi + \theta_c} f(\theta, \phi) \rho_k(\theta - \Phi),$$

where $\Phi = \phi$ for the perfect map and $\rho_0(\theta) = 1$, $\rho_k(\theta) = \cos 2k\theta$ for $k > 0$ and $\rho_k(\theta) = -\sin 2k\theta$ for $k < 0$. In restricted Fourier modes the time evolution of I reads:

$$\tau_w \frac{\partial}{\partial t} \delta \widehat{I}_k(\phi) = -\delta \widehat{I}_k(\phi) + \int_{\phi - \theta_c}^{\phi + \theta_c} \frac{d\theta}{\pi} \rho_k(\theta - \phi) \int \frac{d\phi'}{\pi} h(\phi - \phi') \delta m(\theta, \phi').$$

The restriction is prescribed by the pattern ϕ occurring in the input function on the left hand side. In the activity on the right hand side another pattern, ϕ' , occurs. In order to get closed expressions in the restricted components of I , we need to go to the Fourier dual of the pattern space as well. These Fourier integrals are non-restricted, i.e. calculated over the full half circle $[0, \pi]$. After a straight-forward but quite lengthy analysis, we have found three types of instabilities of the perfect map. First, there is a *rotational instability* induced by a marginally stable mode due to rotational symmetry of the system that only shifts the tuning. Second, we find a *amplitude instability* corresponding to fluctuations that leave the tuning untouched, but send the activity level into saturation. This is the dominant instability if the level of inhibition is too low. Third, there exists a *pattern bias instability* caused by a non-zero average input. The components of the patterns $\mathbf{x}(\phi)$ represent firing rates of LGN-neurons and hence have a positive average activity. The system can always ‘use’ this bias in the LGN patterns to form orientation-independent activity in the cortical hypercolumn, thus unlearning the perfect map.

A bias in the input leading Hebbian learning astray is not a new phenomenon. A well-known remedy is removing the bias of the patterns. Here this leads to

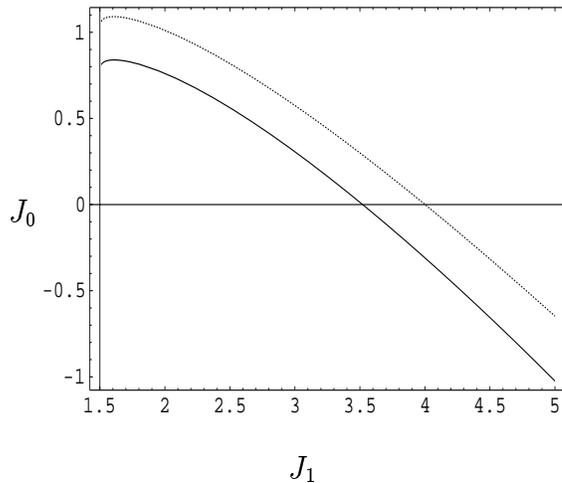


Fig. 1. Stability of perfect map for $h_1 = 0.5$. Below the dotted line the perfect map is stable under neuronal dynamics, above this line the activity is unbounded. Below the solid line the perfect map is also fully stable under modified Hebbian learning. To the left of the line $J_1 = 1.5$ no orientation tuned (stable or unstable) solution exists (as $J_1 + h_1 < 2$).

the modified learning rule:

$$\tau_w \frac{\partial}{\partial t} \mathbf{w}(\theta) = -\mathbf{w}(\theta) + \int \frac{d\phi}{\pi} (\mathbf{x}(\phi) - \langle \mathbf{x} \rangle) \overline{m}(\theta, \phi),$$

where $\langle \cdot \rangle$ is an average over all input angles. Indeed, this modification stabilises the configuration of the hypercolumn as studied by Ben-Yishai et al. The bias-less patterns have no h_0 component in their overlap, but the non-constant overlap is identical to that of the original patterns, so that the new rule is a particular limit of the old rule. Removal of the pattern-bias is here found to be equivalent to constraining the dynamics to the hyperplane $\sum_i w_i(\theta) = 0$, i.e. to keeping a balance between excitation and inhibition. This is more similar to the constraints used in early computer simulations by Von der Malsburg [4]. The overall majority of the input to the cortex can still be excitatory as we ignored all synapses that are fixed on this time-scale.

4 Conclusion

Assuming Hebbian plasticity in the LGN \rightarrow V1 connections we have shown analytically that a stable orientation tuned hypercolumn configuration as proposed by Ben Yishai et al. [2] can emerge in principle. Although this configuration is unstable under plain Hebbian learning with weight decay, we have calculated that removing the average LGN-activity from the learning rule, or fixing the balance between excitatory and inhibitory connections, is enough to ensure

the stability in a large region of parameter space. The main criticism of the model studied here is that no natural images are presented to the network. Although we do not need to specify the patterns very explicitly, we have taken them from a circular manifold in the input space. Further research can be done by extending the stability analysis to less restricted inputs.

Acknowledgements

The authors wish to thank the organisers and sponsors of the EU Advanced Course in Computational Neuroscience 1999 in Trieste, where part of this work was done.

References

- [1] S.-I. Amari. Topographic organization of nerve fields. *Bull. Math. Biol.* 42 (1980) 339–364.
- [2] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky. Theory of orientation tuning in visual cortex. *Prod. Nat. Acad. Sci.*, 92 (1995) 3844–3848.
- [3] D. Hansel and H. Sompolinsky. Modeling feature selectivity in local cortical circuits, in C. Koch and I. Segev, ed., *Methods in Neuronal Modeling*, (MIT Press, 1998) 499–567.
- [4] C. von der Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14 (1973) 85–100.
- [5] G. Sclar and R. D. Freeman. Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Exp. Brain Res.*, 46 (1982) 457–461.