Development of retinotopy and ocular dominance by soft topology-preserving maps and elastic nets: Deriving one class of the models from another

Valery Tereshko
School of Computing, University of Paisley, Paisley PA1 2BE, Scotland
e-mail: valery.tereshko@paisley.co.uk

Abstract

The batch version of soft topology-preserving map producing retinotopy and ocular dominance in visual cortex is proven to be reduced to the elastic net. This verifies numerous results of numerical simulations described in the literature demonstrating similarities of neural patterns produced by lateral and elastic synaptic interactions.

1 Introduction

Neural receptive fields of visual systems are ordered. The projections from retina to optic tectum (in lower vertebrates), and from retina to lateral geniculate nucleus, then to primary visual cortex (in mammals) are topographic. The latter means that neighbouring point in the retina are mapped to neighbouring points in the cortex (tectum). The development of such continuous topographic mapping is called retinotopy. This order guarantees improvement of the recognition abilities and, hence, facilitates the species survival. Indeed, without ordering slight external or internal (neural) noise can results in absolutely unpredictable (and possibly completely wrong) outcome, whereas the ordered receptive field guarantees the recognition of a prototype that is, perhaps, not exactly the same, but very similar to the stimulus.[1] The development of neural receptive fields in a way that they mimic stimulus distribution and become ordered is, thus, biologically meaningful.

Together with retinotopy, ocular dominance and orientation preference are developed.

Mammalian primary visual cortex is naturally binocularly innervated. During development of many, though not all, mammalian species, each part of the visual cortex becomes more densely innervated by one eye and less densely innervated by the other. Eventually, so-called ocular dominance stripes, that are reminiscent of the zebra stripe pattern, are developed. Moreover, exact details of the stripes (their shape, spacing of the pattern, etc) are determined dynamically during development rather than by genetics.

During development the visual cortex cells become largely respond to some preferred orientations. Like ocular dominance, orientation selectivity forms its own pattern: cells with the same orientation preference group to the same domain.

There is strong biological evidence that the same mechanisms underline the formation of both retinotopy, ocular dominance, and orientation selectivity in the visual systems. Different models account for the above phenomena (see reviews [1–3]). However, the only models so far that successfully describe the simultaneous formation of all above processes are based on either the Kohonen approach [4] or the elastic net [5].

Kohonen's self-organizing map utilize lateral interactions to perform a mapping from the stimulus space to the response space with preserving neighbourhood relations. It has been successfully applied to describe the complete development of visual cortex, i.e. simultaneous formation of retinotopy, ocular dominance, and eye and orientation preference [6].

It is known that elastic synaptic interactions can forge a topology as well. The elastic net was first applied to solve the travelling salesman problem (TSP) [5]. This algorithm works like an elastic rubber ring: it gradually drags points on the ring towards the cities and an elastic force keeps neighbouring points close to one another. Another application of elastic synaptic interactions is the preservation of topology in cortical mappings [7–12].

---

[1] I thanks Dr. Yin who pointed my attention to this issue.

Researchers have paid attention to similarities of the neural patterns produced by lateral and elastic synaptic interactions [3, 15–18], but still there is no rigorous justification of such relationships.

We already demonstrated the benefits of using both lateral and elastic interactions for controlling the receptive field patterns [8–13]. In [14], we considered the model utilizing only lateral interactions, which, unlike the elastic ones, are biologically justifiable, and applied it to the problems previously solved only with elastic interactions. Our current aim consists in the proof of equivalence of both types of interactions. For this purpose, we consider development of retinotopy and ocular dominance by the two models. The first model, the batch version of soft topology-preserving map, is based on weak nearest-neighbour lateral interactions. Considering the weight vector of a cortical neuron as a "particle" moving in space-time of the retina and decomposing this particle trajectory over the retinal patterns, we derive the elastic net. The latter model is based on elastic, diffusion-type, interactions. Numerical simulations justify the theoretical results: both models produce similar ocular dominance patterns.

2 Ocular dominance formation

The idea of cortex as a dimension-reducing map from high-dimensional stimulus space to its two-dimensional surface has proved to be fruitful [3, 7]. Performing a cortical mapping induces two conflicting tendencies: (1) the cortical surface should pass through the representative points in stimulus space; (2) the area of the sheet should be kept a minimum. This ensures the formation of smooth receptive fields and, hence, the minimal "wiring" interconnecting the cortical cells, which, in turn, ensures the closeness of the cortical cells representing similar stimuli.

The above mapping can generate ocular dominance with the visual cortex cells varying in their responsiveness to each eye. The order imposed reveals itself in the in the formation of stripe pattern with neigbouring cells sharing a preference to one of the eyes.

Below, two models based on lateral and elastic interactions, respectively, and the relationship between them are analyzed.

2.1 Batch mode of soft topology-preserving map

We consider a one-dimensional net of n stochastic neurons trained by N patterns. The energy of this net, for a given stimulus, is

$$E_i(\mu) = \frac{1}{2}\sum_{j=1}^{n} h_{ij}|\boldsymbol{x}_\mu - \boldsymbol{w}_j|^2, \tag{1}$$

where $\boldsymbol{x}_\mu$ is a given sample pattern, $\boldsymbol{w}_j$ are the weight vectors, and $h(i,j)$ is the neighbourhood function.

Throughout, we consider nearest-neighbour lateral interactions

$$h_{ij} = \begin{cases} 1, & i = j; \\ \gamma, & |i-j| = 1; \\ 0, & |i-j| \geqslant 2, \end{cases} \tag{2}$$

where $0 < \gamma \leqslant 1$.

Instead of the "hard" assignment of Kohonen original algorithm with an unique winner, we assume a "soft" assignment where every i-th neuron is assigned to a given μ-th pattern with a probability $p_i(\mu)$; $\sum_i p_i(\mu) = 1$ [8–14, 19, 20].

The assignment probabilities minimizing free energy of the system (that is a composite of the averaged energy and thermal noise energy) are found to be

$$p_i(\mu) = \frac{\mathrm{e}^{-\beta E_i}}{\sum_{k=1}^{n} \mathrm{e}^{-\beta E_k}}, \tag{3}$$

which gives the minimal free energy [8–14, 19, 20]

$$F(\mu) = -\frac{1}{\beta}\ln\Big(\sum_{i=1}^{n} \mathrm{e}^{-\beta E_i}\Big). \tag{4}$$

Incremental learning strategies are derived through a steepest descent minimization of function (4). The dynamics follows of the free energy gradient, which result in soft topology-preserving mapping [8–14]. Soft mapping is based on soft competition which allows all neurons to adjust their weights with probabilities proportional to their topographic distortion. This makes the weights move more gradually to the presented patterns. The strength of the competition is adjusted by a temperature. The underlying mechanism, deterministic annealing, is derived from statistical physics: it mimics an ordering process during a system's cooling. At high temperatures, the competition is weak and the original energy landscape is smoothed by noise, which helps to eliminate local minima at the beginning of the ordering phase. On reducing the temperature, the competition becomes stronger, the smoothing effect gradually disappears, and the free energy landscape resembles the original one.

In this paper, we consider the batch learning mode when the updating rule is averaged over the set of training patterns before changing the weights.

This gives the following free energy:

$$F = -\frac{1}{\beta N}\sum_{\mu=1}^{N} \ln\Big(\sum_{i=1}^{n} \mathrm{e}^{-\beta E_i}\Big). \qquad (5)$$

Minimization of energy (5) results in the batch version of soft topology-preserving map:

$$\begin{aligned}\Delta \boldsymbol{w}_j &= -\eta \frac{\partial F}{\partial \boldsymbol{w}_j} \\ &= \frac{\eta}{N}\sum_{\mu=1}^{N}\sum_{i=1}^{n} p_i(\mu) h_{ij}(\boldsymbol{x}_\mu - \boldsymbol{w}_j), \qquad (6)\end{aligned}$$

where η is the learning rate.

At low temperatures ($\beta \to \infty$), (6) reduces to the batch mode of the Kohonen map. Goodhill applied the latter model with the special lateral interaction function to modelling the formation of topography and ocular dominance in the visual cortex [1]. The elastic net was found to slightly outperform the Kohonen map [1]. Unlike Kohonen's map, the elastic net algorithm is based on soft competition, with the temperature effectively incorporated into the system [8–13, 21]. Soft competition, as a rule, leads to the development of more precise topography, which can be a reason for better performance of the elastic net. Another limitation of the Kohonen map is the use of a shrinking lateral interaction function, which is not biologically justified [23].

Here, we apply the mapping (6) to modelling the the formation of retinotopy and ocular dominance. Throughout, the training is cyclic with a fixed sequence, i.e. before learning starts a particular order of pattern presentation is fixed. We applied the mapping (6) to model this phenomenon. The simulations are performed for 32 cortical neurons with initial uniform random distribution of the weight vectors within $[-0.0667, 0.0667] \times [-1, 1]$ rectangular (Fig. 1(a)). The stimuli are placed regularly within the two columns of 16 units each at the left and right boundary of the rectangular, which represent left and right "eyes" respectively. The ratio of the separation units between retinae to the separation of neighbouring units within a retina defines the correlation of retinal units, which is ≈ 1. The lateral interactions are weak ($\gamma = 0.03$). The learning rate (with $\eta_0 = 1$) linearly decreases. Let us look at evolution of the weight vector distribution when the inverse temperature increases from $\beta_s = 4$ to different values of β_e in steps of 0.01. For β_e exceeding some threshold but remaining relatively small, the weight vectors become distributed on a line exactly between the left and right eyes (Fig. 1(b)). For intermediate values of β_e, the clusters consisting two weight vectors are formed on this line (Fig. 1(c)). Increasing β_e leads to breaking of the spatial symmetry when the clusters move to either left or right eyes. The latter means the simultaneous formation of retinotopy and ocular dominance (Fig. 1(d)).

2.2 Elastic net

The elastic net is based on elastic, diffusion-type, interactions [5]. Applied to the cortex, elastic synaptic interactions showed their efficiency to preserve the topology [7]. In one approach, the cortex is modelled as a dimension-reducing map from a high-dimensional stimulus space to its two-dimensional surface. Utilizing soft competition mechanism and synaptic interactions, this mapping pulls the cortical sheet towards the stimuli but, at the same time, keeps the cortical area as minimal as possible. Another approach uses elastic nets to model the formation of ocular dominance stripes [24, 25].

Earlier, Simic showed the relationship between the Hopfied network and the elastic net: it derived the latter from Hopfield's objective function for the TSP [21, 22]. Let us show how to derive the elastic net from the batch version of soft topology-preserving map.

Taking the Taylor series expansion (in power of γ) results in

$$\begin{aligned}F = &-\frac{1}{\beta N}\sum_{\mu=1}^{N}\ln\sum_{i=1}^{n}\exp\big(-\frac{\beta}{2}|\boldsymbol{x}_\mu - \boldsymbol{w}_i|^2\big) \\ &+ \frac{\gamma}{2N}\sum_{i=1}^{n}\sum_{\mu=1}^{N} p_i(\mu)\big(|\boldsymbol{x}_\mu - \boldsymbol{w}_{i-1}|^2 \\ &+ |\boldsymbol{x}_\mu - \boldsymbol{w}_{i+1}|^2\big)\ . \qquad (7)\end{aligned}$$

Following Simic's approach, consider the weight vector as a "particle" moving in space-time $\boldsymbol{x}$ and decompose this particle trajectory:

$$\boldsymbol{w}_j = \langle \boldsymbol{x}(j)\rangle = \sum_{\nu=1}^{N} p_j(\nu)\boldsymbol{x}_\nu, \qquad (8)$$

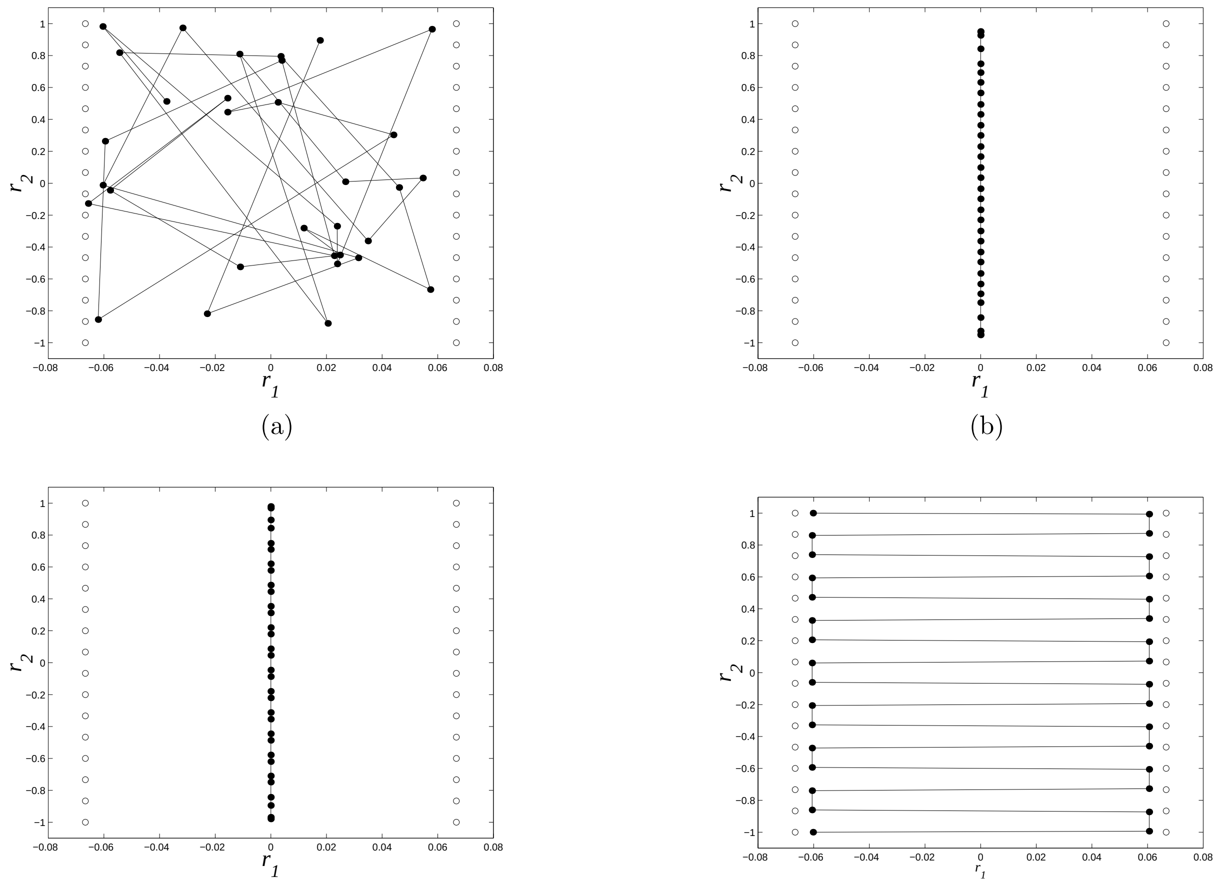
where $\boldsymbol{x}(j)$ and $\langle \boldsymbol{x}(j)\rangle$ are the position and the expected position of the particle at time j respectively.

Applying decomposition (8) to free energy (7) and taking the low temperature limit yield

$$\begin{aligned}F = &-\frac{1}{\beta N}\sum_{\mu=1}^{N}\ln\sum_{i=1}^{n}\exp\big(-\frac{\beta}{2}|\boldsymbol{x}_\mu - \boldsymbol{w}_i|^2\big) \\ &+ \frac{\gamma}{N}\sum_{i=1}^{n}|\boldsymbol{w}_{i+1} - \boldsymbol{w}_i|^2\ . \qquad (9)\end{aligned}$$

Minimization of free energy function (9) results in the elastic net algorithm:

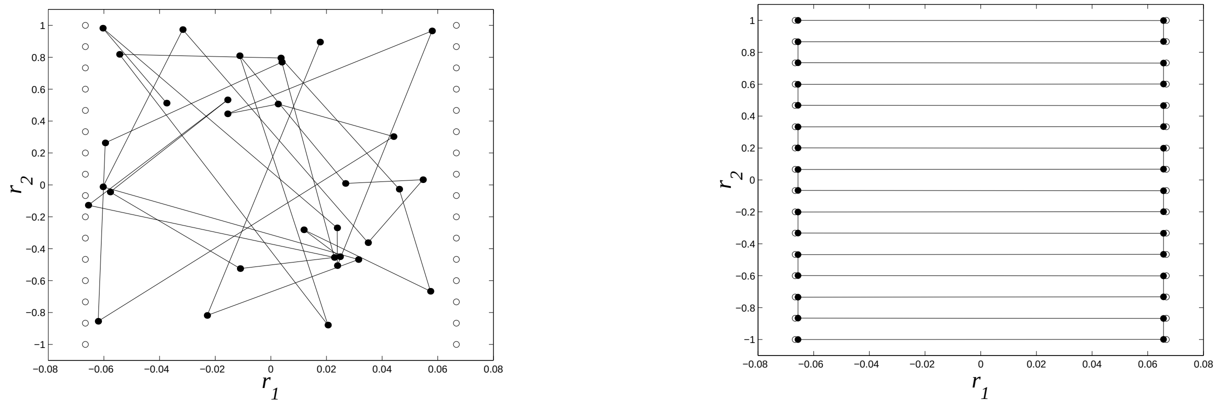
$$\begin{aligned}\Delta \boldsymbol{w}_j = &-\eta\frac{\partial F}{\partial \boldsymbol{w}_j} = \frac{\eta}{N}\Big(\sum_{\mu=1}^{N}\tilde{p}_j(\mu)(\boldsymbol{x}^\mu - \boldsymbol{w}_j) \\ &+ 2\gamma(\boldsymbol{w}_{j+1} - 2\boldsymbol{w}_j + \boldsymbol{w}_{j-1})\Big)\ , \qquad (10)\end{aligned}$$



(a) (b)

(c) (d)

Figure 1: Weight vector distribution of the cortical chain: (a) initial, (b) – (d) after applying the mapping (6) with $\gamma = 0.03$ and (b) $\beta_e = 150$, (c) $\beta_e = 250$, and (d) $\beta_e = 1000$ (see details in the text). Stimuli and weight vectors are marked by open and filled circles, respectively.



(a) (b)

Figure 2: Weight vector distribution of the cortical chain: (a) initial, (b) after applying the mapping (10) with $\gamma = 0.03$ and $\beta_e = 1000$ (see details in the text). Stimuli and weight vectors are marked by open and filled circles, respectively.

where

$$\tilde{p}_j(\mu) = \frac{\exp(-\frac{\beta}{2}|\boldsymbol{x}^\mu - \boldsymbol{w}_j|^2)}{\sum_{k=1}^{n} \exp(-\frac{\beta}{2}|\boldsymbol{x}^\mu - \boldsymbol{w}_k|^2)} \qquad (11)$$

is a reduction of $p_j(\mu)$ to the case of no lateral interactions. Defining $\beta \equiv \frac{1}{\sigma^2}$ with the Gaussian distribution width σ, energy (9) takes the exact form of the Durbin-Willshow elastic net energy [5]. Shrinking the distribution width is, thus, equivalent to reducing the system temperature.

For elastic net (10), we performed the same numerical simulations that were done for soft topology-preserving map (6). Our theoretical findings are verified by numerical results confirming the similarity of both modules. Comparison of patterns in Figs. 1(d) and 2(b) are clearly demonstrates the latter.

3 Conclusions

Utilizing soft competition instead of a hard one gives an opportunity (i) to derive the system energy function, and (ii) to describe the learning process in terms of deformation of the energy landscape keeping the lateral interaction range fixed. Such a deformation of the landscape may be caused by a compound action of interactions of the retinocortical system cells, diffusion of free molecules, and neuronal noise. This gives an alternative to widely used but not justified learning based on shrinking the lateral interaction range. In such a way, by utilizing fixed nearest-neighbour lateral interactions, soft competition, and deterministic annealing, the topology is forged in the system.

Remarkably, the elastic net can be derived from the batch version of soft topology-preserving map. The nearest-neighbour lateral interactions are transformed to the elastic ones, which proves their similarity for forging the cortex topology. In turn, the Gaussian variance of elastic net appears to be the temperature of the soft topology-preserving map. The latter elucidates indirect incorporation of soft competition and deterministic annealing into the elastic nets.

The one-dimensional topology-preserving map can be generalized to high-dimensional one and reduced to high-dimensional elastic net respectively. Unlike the original one-dimensional elastic net aimed to pass through the localized representative points (“cities” on a plane), its high-dimensional version can cope with the distributed objects in the stimulus space.

References

[1] G.J. Goodhill, Correlations, competition, and optimality: Modelling the development of topography and ocular dominance, *Cognitive Science Research Paper 226*, University of Succex, 1992.

[2] E. Erwin, K. Obermayer, and K. Schulten, Models of orientation and ocular dominance in the visual cortex: A critical comparison, *Neural Computation*, vol. 7, p. 425-468, 1995.

[3] N.V. Swindale, The Development of Topography in the Visual Cortex: a Review of Models, *Network: Computation in Neural Systems*, vol. 7, p. 161-247, 1996.

[4] T. Kohonen, Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, vol. 43, p. 59-69, 1982.

[5] R. Durbin and D. Willshaw, An analogue approach to the travelling salesman problem using an elastic net method, *Nature*, vol. 326, p. 689-691, 1987.

[6] K. Obermayer, G.G. Blasden, and K. Schulten, Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps, *Physical Review A*, vol. 45, p. 7568-7589, 1992.

[7] R. Durbin and G. Mitchison, A dimension reduction framework for understanding cortical maps, *Nature*, vol. 343, p. 644-647, 1990.

[8] V. Tereshko, Topology-preserving elastic nets, in J. Mira, A. Prieto (Eds.) *Connectionist Models of Neurons, Learning Processes and Artificial Intelligence*, Lecture Notes in Computer Science 2084, Springer, Berlin, 2001, p. 551-557.

[9] V. Tereshko, Topology-preserving neural models with lateral and elastic interactions, *Proc. ICNNAI-2001*, Minsk, 2001, p. 24-31.

[10] V. Tereshko, Topological ordering in neural nets with weak nearest neighbour interactions, *Proc. IJCNN-2001*, Washington, DC, 2001, vol. 1, p. 296-301.

[11] V. Tereshko and N.M. Allinson, Theory of topology-preserving elastic nets, W. Klonowski (Ed.), *Attractors, Signals and Synergetics*, PABS Science Publ. 2002, p. 215 - 221.

[12] V. Tereshko and N.M. Allinson, Combining lateral and elastic interactions: Topology-preserving elastic nets, *Neural Processing Letters*, vol. 15, 2002, p. 213-223.

[13] V. Tereshko and N.M. Allinson, Common framework for "topographic" and "elastic" computations, in D. S. Broomhead, E. A. Luchinskaya, P. V. E. McClintock, T. Mullin (Eds.) *Stochaos: Stochastic and Chaotic Dynamics in the Lake*, American Institute of Physics, Woodbury, 2000, p. 124-129.

[14] V. Tereshko, Modelling topology preservation in a cortex by controlled deformation of its energy landscape, *Neurocomputing*, vol. 44-46C, p. 667-672, 2002.

[15] J. Hertz, A. Krogh, and R.G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Reading, 1991.

[16] R. Folk, A. Kartashov, A simple elastic model for self-organizing topological mappings, *Network: Computation in Neural Systems*, vol. 5, p. 369-387, 1994.

[17] Y. Wong, A Comparative study of the kohonen self-organizing map and the elastic net, in S. J. Hanson, T. Petsche, M. Kearns, R. L. Rivest (Eds.) *Computational Learning Theory and Natural Learning Systems*, vol. 2, MIT, Cambridge, 1994, p. 401-413.

[18] G. Mitchison, A Type of duality between self-organizing maps and minimal wiring, *Neural Computation*, vol. 7, p. 25-35, 1995.

[19] T. Graepel, M. Burger, and K. Obermayer, Phase transitions in stochastic self-organizing maps, *Physical Review E*, vol. 56, p. 3876-3890, 1997.

[20] T. Heskes, Energy functions for self-organizing maps, in E. Oja, S. Kaski (Eds.) *Kohonen Maps*, Elsevier, Amsterdam, 1999, p. 303-315.

[21] P.D. Simic, Statistical mechanics as the underlying theory of "elastic" and "neural" optimisations, *Network: Computation in Neural Systems*, vol. 1, p. 89-103, 1990.

[22] J. van den Berg, Neural relaxation dynamics: Mathematics and physics of recurrent neural networks with applications in the field of combinatorial optimization, *PhD Thesis*, Erasmus University Rotterdam, 1996.

[23] G.J. Goodhill, Topography and ocular dominance: A model exploring positive correlations. *Biological Cybernetics*, vol. 69, p.109-118, 1993.

[24] G.J. Goodhill and D.J. Willshaw, Application of the elastic net algorithm to the formation of ocular dominance stripes, *Network: Computation in Neural Systems*, vol. 1, p. 41-59, 1990.

[25] G.J. Goodhill and D.J. Willshaw, Elastic net model of ocular dominance: Overall stripe pattern and monocular deprivation, *Neural Computation*, vol. 6, p. 615-621, 1994.