A comparative evaluation of the Generative Topographic Mapping and the Elastic Net for the formation of Ocular Dominance stripes

Dror Cohen, Andrew P. Papliński
Clayton School of Information Technology
Monash University
Email: {Dror.Cohen, Andrew.Paplinski}@monash.edu

***Abstract*—In this paper we compare the self organising capabilities of the Generative Topographic Map (GTM) [1] and Elastic Net (EN) [2]. We analytically compare the two algorithms and examine the different ways in which they preserve topography by considering their respective 'state space trajectories'. We present simulations that demonstrate the differences between the two algorithms. We conclude by using the GTM to simulate the formation of Ocular Dominance (OD) stripes and compare against earlier simulations using the EN. Our findings indicate that the GTM produces patterns with some of the required characteristics and match results obtained with the EN to a degree.**

***Index Terms*—Generative Topographic Mapping, Self-organisation, Elastic Net, Ocular Dominance**

I. Introduction

In this paper we compare the self organising capabilities of the Generative Topographic Map (GTM) [1] and Elastic Net (EN) [2]. We present analytical and experimental comparisons and investigate the way topography is preserved by considering the algorithms in 'state space'. We present simulations that highlight the differences between the two algorithms and conclude by using the GTM to simulate the formation of Ocular Dominance (OD) stripes.

Today topography preserving latent space models are widely used for data visualisation and dimensionality reduction in a variety of fields (e.g. [3], [4], [5]). The best known such model is the Self-Organizing Map [6]. Much of the current research is focused on formulating and improving these models through probabilistic treatments, in line with current trends in Machine Learning [7], [3]. Both the GTM and EN are examples of such probabilistic models and are often referred to as probabilistic alternatives to the SOM. One prominent use of the EN is to model the formation of OD stripes, which are formed as neurons in early visual processing associate more strongly with one of the two eyes. This results in a stripe-like pattern segregating neurons with different ocular preferences [8], [9], [10]. The use of self organising models to simulate this phenomena is often motivated under a 'wire length minimisation' approach, where the brain is alleged to develop in a way that minimises the length of neuronal connections [11], [12]. From a functional perspective, it has been suggested that the brain may utilise such an ordered representation in a hierarchical fashion to reduce the dimensionality of the input space [13].

The EN and SOM match the experimentally observed OD patterns very closely, motivating the idea that the Hebbian type learning which both algorithms employ underlies the formation of the observed pattern [14], [9]. It is therefore of interest to determine whether other such models are able to produce these patterns. The model of interest to us is the GTM as the way in which it preserves topography offers biological interpretation in the form of a feed-forward neural network (I-B). Such feed-forward models have been used in earlier attempts to model OD stripes [10] and remain a successful model and active area of research in modelling of early visual processing [15].

Further, there is a large body of work in Computational Neuroscience that has successfully employed probabilistic treatments [16], [17], [18]. Probably the most extensive such example is the recently proposed free-energy principle [19], [20], [21] which is centred around a type of probabilistic inference. Therefore, explaining the OD stripes formation process using a probabilistic model may allow us to generalise some of these treatments.

A. Elastic Net

The Elastic Net (EN) was first developed as an analogue approach to the Travelling Salesman Problem [2].

In the EN, nodes (corresponding to neurons or neuronal populations) are evenly spaced on a latent space (usually 2D) lattice. Each node m has a position vector $v_m \in \mathbb{R}^2$ and a weight vector $w_m \in \mathbb{R}^D$. A set of observation $x_n \in \mathbb{R}^D$ $X = \{x_1, x_2, ..., x_N\}$ is then introduced for which, the weights are updated according to the following rule:

$$\Delta w_m = \sum_n \rho_{nm}(w_m - x_n) + \alpha k \sum_j |w_m - w_{j\in\Lambda_m}|^2 \quad (1)$$

Here $w_{j\in\Lambda_m}$ represents the weights of a node j, and Λ_m contains all the nodes in the 'lattice neighbourhood' of node m, and α and k are constants. Typically, only nearest neighbour interaction (directly South, North, East and West) is assumed. The terms ρ_{nm} are the 'responsibilities' given by

$$\rho_{nm} = \frac{\Theta(x_n|w_m, k)}{\sum_s \Theta(x_n|w_s, k)} \quad (2)$$

with $\Theta(x_n|w_m, k) = \exp(\frac{-|x_n - w_m|^2}{2k^2})$.

An EM type algorithm [22] is employed where at each iteration the responsibilities (2) are calculated, followed by the update equation (1) and consequent recalculation of the responsibilities. Throughout this process the constant k is gradually reduced so that the weight of the 'continuity' term (second term in (1)) compared to the 'coverage' term (first term in (1)) is reduced.

One advantage of the EN over the SOM is that it can be seen to preform a gradient decent on the following energy function:

$$\log p(X|W,k) = \\ -k\sum_n \log \sum_m N(x_n|w_m,k) + \frac{1}{2}\alpha \sum_m \sum_j |w_m - w_{j\in\Lambda_i}|^2 \quad (3)$$

where $N(x_n|w_m,k)$ is the normal distribution with mean w_m and variance k, X is a $D\times N$ matrix with columns given by $\{x_1,x_2,...,x_N\}$ and W is a $D\times M$ matrix with columns given by $\{w_1,w_2,...,w_M\}$. In a probabilistic formulation this energy function can be interpreted as *maximum a posteriori* estimate for a Gaussian Mixture Model (GMM) with a Gaussian Process prior over the weights:

$$p(x_n|W,k) = \\ \frac{1}{M}\sum_m N(x_n|w_m,k)\exp(-\frac{1}{\alpha k}\sum_m\sum_j |w_m - w_{j\in\Lambda_i}|^2) \quad (4)$$

Detailed analysis of the Elastic Net and the properties of this specific prior can be found in [23], [24].

B. The Generative Topographic Mapping

Consider a latent space consisting of a square lattice with M nodes evenly spaced on it. As in the EN, each node m has a position $v_m \in \mathbb{R}^2$ and a weight vector $w_m \in \mathbb{R}^D$. In the GTM the energy function that is being minimised corresponds to Maximum Likelihood on a constrained GMM:

$$\log p(X|W) = \sum_n \log \frac{1}{M}\sum_m N(x_n|w_m,\beta^{-1}) \quad (5)$$

where β are the precisions. In the original GTM [1] the constraint takes the form of a Radial Basis Function (RBF) network $\boldsymbol{\phi}(v) = \{\phi_1(v),\phi_2(v),\ldots,\phi_K(v)\}^T$, with the RBF centres evenly spaced throughout the latent space. That is,

$$w_m = g(\Omega,v_m) = \Omega\boldsymbol{\phi}_m$$
$$\boldsymbol{\phi}_m = \{\phi_1(v_m),\phi_2(v_m),...,\phi_K(v_m)\}^T$$

The update equations are obtained by differentiating (5) w.r.t the weight matrix Ω and the precisions β:

$$\Phi^T G\Phi\Omega_{new}^T = \Phi^T RX \quad (6)$$

$$\frac{1}{\beta_{new}} = \frac{1}{ND}\sum_n^N\sum_m^M \rho_{nm}|\Omega_{new}\boldsymbol{\phi}_m - x_n|^2 \quad (7)$$

where Φ is a $K\times M$ matrix with columns given by $\boldsymbol{\phi_1},\boldsymbol{\phi_2},\ldots,\boldsymbol{\phi_m}$, R is a $K\times N$ matrix with elements given by the nodes' responsibilities:

$$R_{nm} = \rho_{nm} = \frac{N(x_n|w_m,\beta^{-1})}{\sum_m N(x_n|w_m,\beta^{-1})} \quad (8)$$

and G is a $K\times K$ diagonal matrix with elements $G_{kk} = \sum_n \rho_{nk}$. Here too an EM type strategy is employed where the responsibilities are evaluated in (8) followed by the update equations (6) and (7). Further technical details can be found in [1], [25].

In the GTM the mapping $w_m = g(\Omega,v_m) = \Omega\boldsymbol{\phi_m}$ determines the relationship between the weights in W, that is, the way in which topography is preserved. When the RBF functions are chosen as isotropic Gaussians $\phi_k(v_m) = \exp(-\lambda|v_m - c_k|^2)$ (with c_k denoting the position of the kth RBF centre on the lattice) the parameters that control the mapping are the number of RBF centres K and the spread parameter λ.

This mapping can be interpreted as a transformation between two neuronal lattices where the efficacy of connections decreases as connection length increases, as demonstrated in Fig. 1.

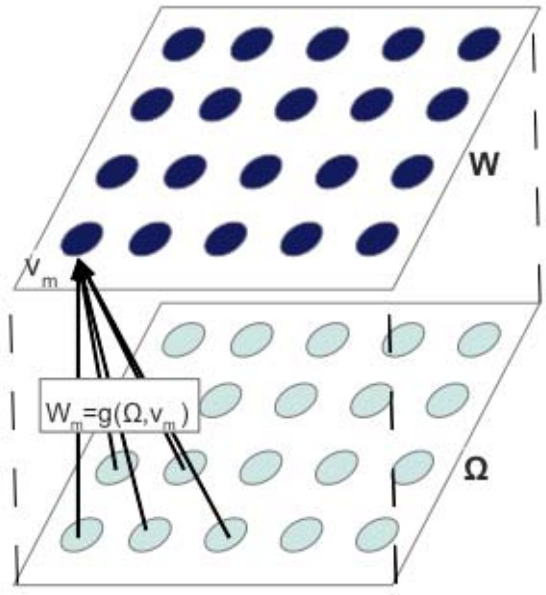

Fig. 1. Interpreting the mapping in the GTM as a mapping between neuronal lattices when the efficacy of connection decreases with connection length. The bottom lattice represents the underlying weight matrix Ω that is mapped to the top lattice, corresponding to W. Subset of connections shown.

A quadratic regularisation term of the form

$$(\frac{\alpha}{2\pi})^{(D\cdot M)/2}\exp(-\frac{\alpha}{2}\sum_m |w_m|^2)$$

can be added to the objective function (5) to control the smoothness of the mapping. In a probabilistic setting, this regularisation term can also be viewed as a prior over the weights. However, with the inclusion of this term the simple feed-forward interoperability is complicated and so we do not consider this extension.

In fact, it is possible to abandon the 'hard mapping' altogether and achieve topography preservation through the consideration of an appropriate prior over the weights alone. This is known as the Gaussian Process GTM [25]. Indeed, for

an identical choice of prior the Gaussian Process GTM and EN have identical objective functions.

We are unaware of any definitive studies preferring either the original or Gaussian Process GTM. Also, since its original conception, improvements to both the original and the Gaussian Process GTMs have been proposed. Some earlier ones can be found in [25], while more recent ones include optimising through Deterministic Annealing [26], employing a Variational treatment [27] and using the 'kernel trick' [28] .

II. Comparing objective functions and mechanism for topography preservation

A. Objective functions

By equating the EN and GTM objective functions (4), (5) it is possible to obtain a mapping $g(\Omega, v_m)$ for which the GTM and EN are identical, that is:

$$\text{objective function GTM} = \text{objective function Elastic Net}$$

$$\begin{aligned}
&\sum_n \sum_m a_{nm} \left\{ \log(\frac{\beta}{2\pi})^{D/2} + |x_n - g(\Omega, v_m)|^2 \right\} \\
= &\sum_n \sum_m b_{nm} \left\{ \log(\frac{k}{2\pi})^{D/2} + |x_n - w_m|^2 \right\} \\
&+ \alpha k \sum_m \sum_j |w_m - w_{j \in \Lambda_m}|^2 \qquad (9)
\end{aligned}$$

where a_{nm} and b_{nm} correspond to the responsibilities for the GTM (8) and EN (2) respectively. To obtain (9) we have made use of the Hidden Variable derivation of the EM (e.g. Chapter 9, [29]). The expression can be further simplified if only the update equation for the weights are considered.

$$\begin{aligned}
&\sum_n \sum_m a_{nm} |x_n - g(\Omega, v_m)|^2 \\
= &\sum_n \sum_m b_{nm} |x_n - w_m|^2 + \alpha k \sum_j |w_m - w_{j \in \Lambda_m}|^2; \\
&\sum_n \sum_m a_{nm} (x_n \cdot x_n - 2x_n \cdot g(\Omega, v_m) \\
&+ g(\Omega, v_m) \cdot g(\Omega, v_m) \\
= &\sum_n \sum_m b_{nm} (x_n \cdot x_n - 2x_n \cdot w_m + w_m \cdot w_m) \\
&+ \alpha k \sum_j (w_m \cdot w_m - 2w_m . w_j + w_j \cdot w_j); \\
&\sum_n \sum_m a_{nm} (-2x_n \cdot g(\Omega, v_m) + g(\Omega, v_m) \cdot g(\Omega, v_m)) \\
= &\sum_n \sum_m b_{nm} (-2x_n \cdot w_m + w_m \cdot w_m) \\
&+ \alpha k \sum_j (w_m \cdot w_m - 2w_m \cdot w_j + w_j \cdot w_j)
\end{aligned}$$

To further simplify this expression we make the assumption that the weights form unity vectors ($w_j \cdot w_j = 1$), which can be enforced by projecting the weights to the unity hypersphere.

$$\begin{aligned}
&\sum_n \sum_m a_{nm} (2x_n \cdot g(\Omega, v_m)) \\
= &\sum_n \sum_m b_{nm} (2x_n \cdot w_m) + \alpha k \sum_j (2w_m \cdot w_j) + C
\end{aligned}$$

where the constants have been grouped into C. One solution to this can be obtained by equating the summed expressions which then form a system of linear equations

$$\begin{aligned}
&a_{nm} (2x_n \cdot g(\Omega, v_m)) \\
= &b_{nm} (2x_n \cdot w_m) + \alpha k \sum_j (2w_m \cdot w_j) + C, \quad \forall m, n
\end{aligned}$$

The above treatment demonstrates that there exist a mapping for which the two are equal, although the mapping will not directly depend on the lattice positions and unlikely have a simple form such as that typically used in the GTM. Further, the dependency on the data cannot be removed.

B. Mechanism for Topography Preservation

The difference in the ways with which the EN and GTM achieve topography can clearly be demonstrated by considering a node's 'state space trajectory' throughout the learning process. In this context the 'state' of a node is the set of its attributes. For the EN, the attributes are the node's position and weight. For the GTM the precision may also be considered an attribute though this would have limited impact as it is common across all nodes.

Fig. 2 depicts the 'state space' trajectory of a node through the learning process, marked by iterations for the EN and GTM.

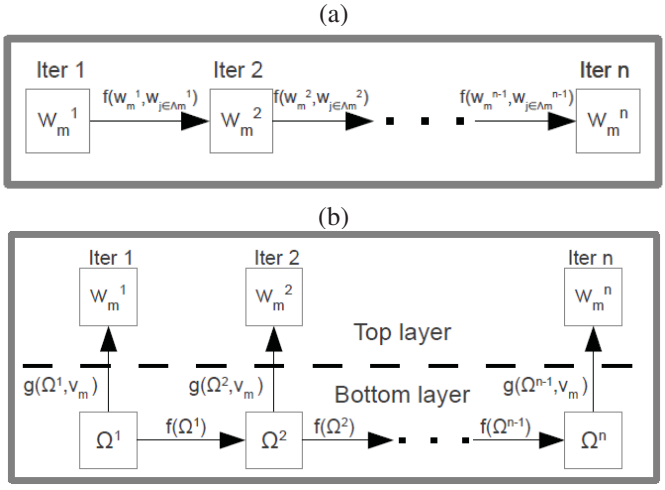

Fig. 2. Weight space trajectories for the EN (a) and for the GTM (b). In the EN neighbouring nodes converge towards the same weight. In the GTM the mapping ensures that two nearby nodes have similar weights, throughout

At each iteration the node's 'state' evolves according to a predefined (deterministic) 'transition function' f_m. For both the EN and the GTM this function is complex due to the responsibilities. In the EN the function also considers nearby nodes resulting in nearby nodes converging to similar states.

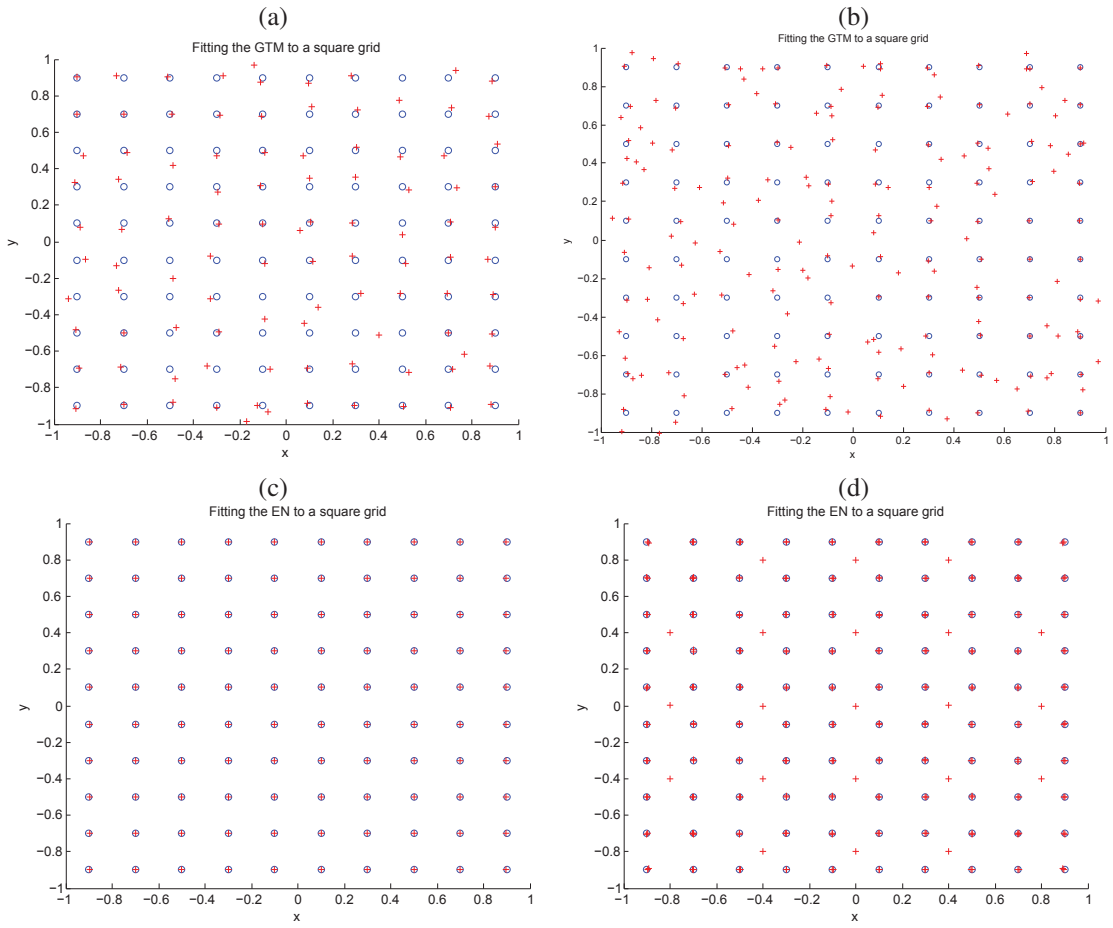

Fig. 3. Using the GTM and EN to fit a 2D latent space to a 2D grid: (a) using the GTM with 100 nodes, (b) using 225; (c) using the EN with 100 nodes, (d) 225 nodes.

In the GTM, nearby nodes always have similar states due to the mapping g. Note that since the data is processed in batch mode it can be incorporated into the transition function and does not need to be represented explicitly.

Positing topography preservation in this way makes clear some of the differences between the GTM and EN. For example, consider three neighbouring nodes with positions v_a, v_b and v_c such that v_b is between v_a and v_c. In the EN, it is possible for $w_a = w_b = w_c$. This can readily be verified by noting that for (4) with $\alpha \to 0$, one obtains an ordinary mixture model, for which any node has no notion of its latent space position. In the GTM, $g(\Omega, v_a) = w_a = w_c = g(\Omega, v_c)$ will typically mean that $w_c, w_a, \neq w_b$, as the mapping will vary smoothly from w_a to w_c. This 'squeezing' property may be overcome by increasing the number of weights in Ω, increasing the precision β and or by some non-arbitrary initialisation of the parameters, which is usually employed when using the GTM [1].

III. Experiments

The simulations in this section were preformed using MATLAB [30]. An implementation of the EN can be found in [31]. An implementation of the GTM can be found in [32], though for flexibility considerations we use a custom implementation.

A. Regular Grid and Clustering

We first evaluate the EN and GTM by fitting a 2D latent space to a regular 10×10 grid as displayed in Fig. 3. For both methods we initialise the weights randomly around zero. It can been seen that both the GTM and EN approximate the grid, however, in the GTM the nodes do not settle on the data points as closely as they do with the EN. Further, the 'squeezing' property typically prevents nodes from collapsing on to each other, which is readily demonstrated by increasing the number of nodes (Fig. 3 a, b). In the EN it is possible for two (or more) nodes to coincide, which we demonstrate by increasing the number of nodes (Fig. 3 c, d)).

Applying the EN on the Oil Phase data (Fig. 4, means shown) we were unable to get results comparable to that of the GTM (shown in [1]). However, the results reported in [1] were obtained using the regularised GTM with PCA based initialisation. Using the non-regularised GTM with random initialisation we were unable to reproduce the reported results. On other simulations with artificially generated clustered data (not shown) the two methods were comparable.

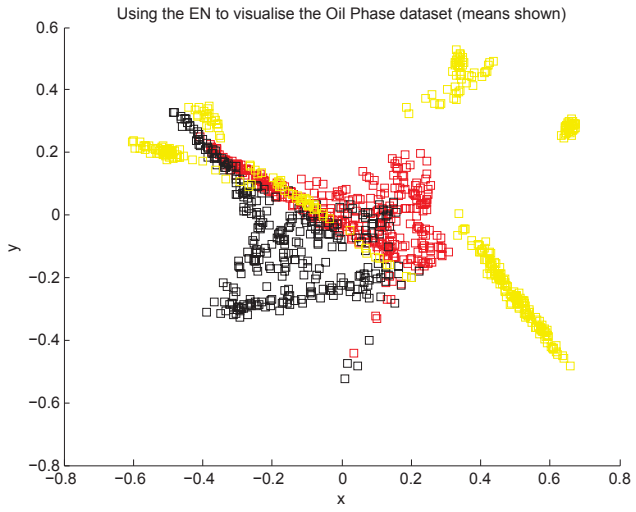

Fig. 4. Using the EN to visualise the Oil Phase data. The different flows are marked by different colours

B. Ocular Dominance Column Formation using the GTM

In this section we apply the GTM to the formation of Ocular Dominance stripes. Throughout, we will compare with the EN as in [33], where it was first applied to this problem. We begin by considering a simplified 1D case where each retina is composed of an array of N points (cells). Each cell is associated with a marker $[l, -l]$ which indicates to which of the two retinas it belongs to. The distance between neighbouring cells in each retina is denoted by d. The simplified 'cortex' consists of a 'string' with M latent space points arranged along it. This arrangement for $\{N, M, d, l\} = \{20, 40, 0.05, 0.1\}$ is depicted in Fig. 5.

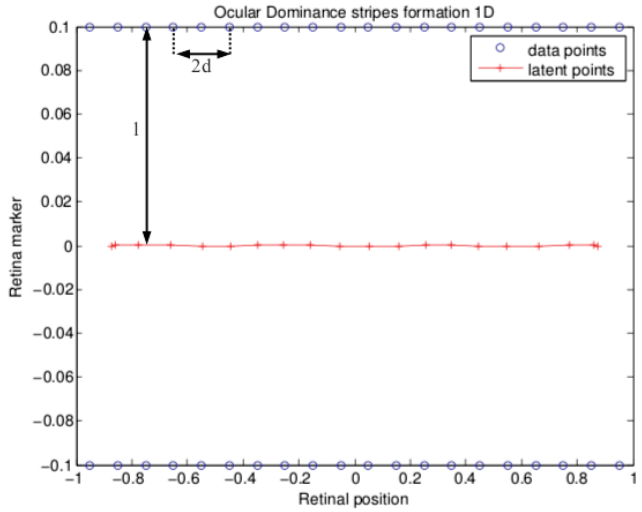

Fig. 5. Fitting a simplified 1D 'cortex' (red) to two 1D retinas (blue)

To allow for flexibility we will use the same number of RBF centres as nodes. The weights are initialised with a small horizontal bias (as in [33]) to correspond to the initial retinotopic arrangement observed and the precision is initially set to some small value. The nodes ocular preference is initialised randomly around zero.

1) Fixed spread: Fig. 6 depicts the resulting patterns for four different values of λ. As expected, if the spread is too great ($\lambda \leq 20$) then the string is too rigid and cannot approximate the data well. Conversely, if the spread is too small (large $\lambda \geq 50$) then the string is too flexible and the retinotopic (horizontal) order is compromised.

We have observed that there is a critical value λ_c at which the string is flexible enough to form a stripe like pattern. For this arrangement it is around $\lambda_c \sim 24.3$ but does exhibit some susceptibility to the initialisation of the weights and precisions. This is typical of the EM and likely amplified by the complex relationship between the spread parameter (λ) and the precisions (β). This value is also proportional to N and d (which effectively control the length of the retina) and inversely proportional to the value l. This is to be expected considering the respective 'forces' a given node experiences (horizontal and vertical pulls) as these values are changed.

The formation process begins with the string first stretching horizontally to lie equidistant between the two arrays. Small fluctuations due to the random initialisation of the ocular preference result in the string adopting a sinusoid like pattern. If the string is sufficiently flexible ($\lambda \geq \lambda_c$) the string undergoes a rapid vertical expansion (formation of stripes), beginning at the ends of the rope (due to boundary effects). If the string is too rigid stripes do not form. This process bears some similarity to that reported in [33].

The number of peaks can be thought of as a measure of stripe width which increases (number of peaks decreases) as the ratio l/d is increased. This can only be reliably reproduced as long as $\lambda \sim \lambda_c$. For $\lambda \gg \lambda_c$, the string twists and the number of peaks cannot be trivially determined. Further, the pattern is no longer homogeneous in this case (see Fig.6) and so the relationship to stripe width is effected. Fig.7 depicts the effect of increasing this ratio. Note that this also changes the value λ_c.

2) Adapting the spread: Solutions akin to that of the EN can be obtained if the mapping is adapted by setting $\lambda_{new} = k\lambda_{old}$ after each iteration, as seen in Fig. 8. Unfortunately, by doing so the GTM's log-likelihood can no longer be guaranteed to increase after each iteration. In principle, the more flexible the mapping the more independently the nodes can move and the more likely it is that the system will converge to a high likelihood state. In practice, we can report that as long as the rate of increase is small enough the log-likelihood predominantly increases, though periods of small decreases do occur. Increasing the value too quickly may result in nonsensical solutions.

We were unable to determine a value for k that ensures ordered solutions consistently, and for all values of l, d and m. Unlike the EN, solutions obtained by a adapting the spread often featured incomplete mappings and areas where twisting has occurred, as shown in Fig. 8.

We also experimented with other cooling schemes such as

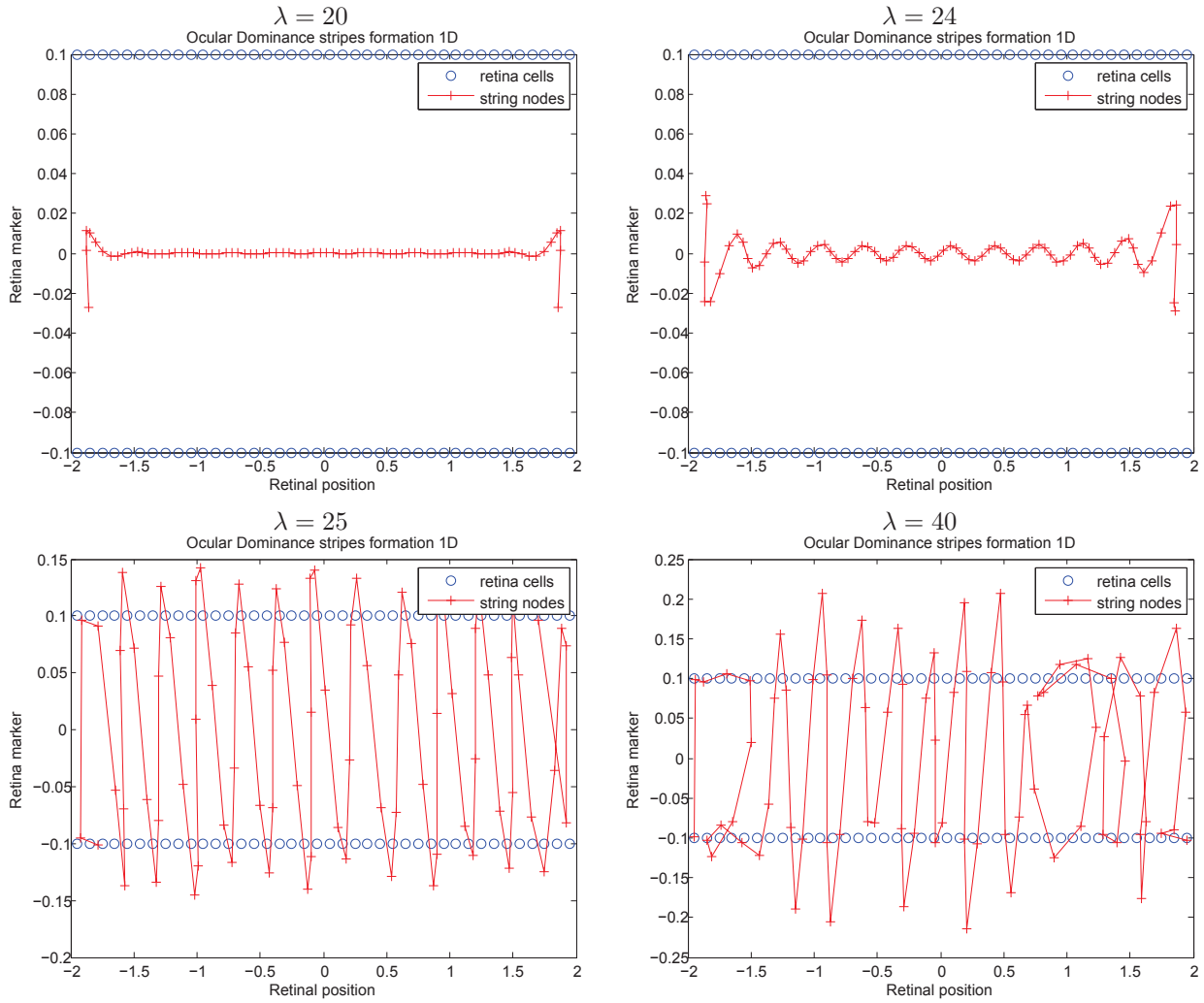

Fig. 6. 1D Stripe formation for a fixed value of $\lambda = 20, 24, 25, 40$. Other parameter values are fixed at $d = 0.1, l = 0.1, M = 80, N = 40$. For small values of λ the manifold is too rigid and for high values the manifold twists and loses the continuous retinotopic property

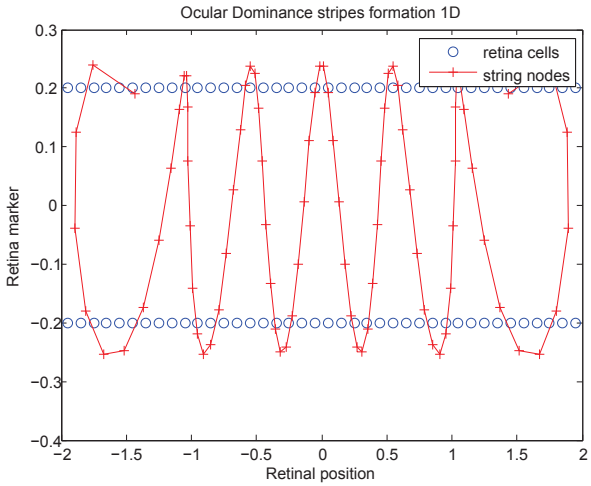

Fig. 7. The ratio l/d effects the stripe width. $d = 0.1, l = 0.2, M = 80, N = 40, \lambda = 5$

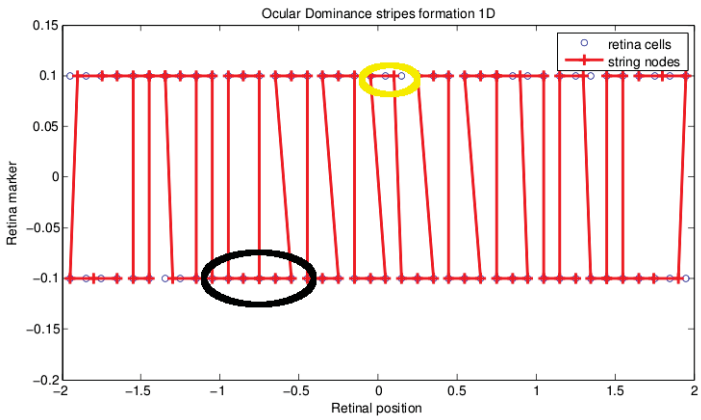

Fig. 8. Adapting the spread value during learning results in solution akin to those obtainable by the EN. The area circled in black shows one region where twisting has occurred. The area circled in yellow shows a data point with no node assigned to it (incomplete mapping)

increasing both the precision and spread iteratively ($\beta_{new} = \alpha\lambda_{new}$) and tying the spread to the precisions ($\lambda_{new} = \frac{1}{\alpha}\beta_{new}$ where β_{new} is determined according to (7)). Both methods were able to produce ordered results but not consistently or robustly.

3) 2D simulation: The 2D case is a straightforward generalisation of the 1D case where the retinas and cortex are modelled as 2D sheets. We have observed no outstanding features for this case. Simulations showing the effect of changing the ratio l/d are shown in Fig. 9

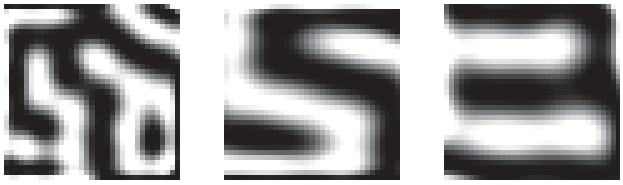

Fig. 9. 2D simulations of OD stripe formation using the GTM. Each retina was modelled as a (20×20) grid. The cortex was modelled as a (35×35) grid. Parameter values from left to right are $\{\lambda = 1, l = 0.1, d = 0.1\}$, $\{\lambda = 0.9, l = 0.15, d = 0.1\}$, $\{\lambda = 0.8, l = 0.2, d = 0.1\}$

IV. Conclusion

We have compared the GTM and EN and demonstrated that the GTM can be interpreted in terms of mappings in between neuronal layers. Comparison of their respective objective functions demonstrated that the two algorithms differ for the usual choice of mapping in the GTM. We examined the different methods the algorithms employ to preserve topography and identified some properties that are unique to each. The simulations preformed confirmed the preceding analysis in that the GTM favours more gradually changing solutions and is less likely to have nodes settling on data points. This is due to the hard mapping in the GTM which does not offer the same flexibility as that obtainable with the EN.

Application of the GTM to model the formation of OD stripes demonstrated that the algorithm produces striped patterns related to those observed. The patterns produced by the GTM and EN have some key differences. Firstly, in the GTM the ocular preferences shifts more gradually than in the EN. This can be seen by observing the gray areas in Fig. 9 which depicts areas of neurons with intermediate ocular preference. Another difference is the stripe width which is typically broader for the GTM.

There are several directions which can be explored further. One straightforward observation is that the EN may be an undervalued tool for clustering and visualisation. The prior used in the EN is also simpler (and sparser) than the one typically used in the Gaussian Process GTM [25], [27]. The use of the EN can be further motivated by noting that effective solutions using Cholesky factorisation instead of gradient decent are possible [24]. A comparison of the Gaussian Process GTM and the EN may help further understand the characteristics of this interesting prior.

For biologically inspired modelling, one may want to maintain the interoperability of the original GTM. In this case, it seems like a principled way of adapting the spread is required to produce results comparable to the EN. This undertaking should also include an evaluation of different mappings (as well as the use of different mappings for different dimensions), a notable exclusion in GTM research. Other directions can include incorporating more complex covariance forms [5], [34] or using one of the recently suggested improvements [27], [28], [26].

References

[1] C. M. Bishop, M. Svensen, and C. K. I. Williams, "Gtm: The generative topographic mapping," *Neural Computation*, vol. 10, no. 1, pp. 215–234, 1998. [Online]. Available: http://www.mitpressjournals.org/doi/abs/10.1162/089976698300017953

[2] R. Durbin and D. Willshaw, "An analogue approach to the travelling salesman problem using an elastic net method," *Nature*, vol. 326, no. 6114, pp. 689–691, 1987. [Online]. Available: http://math.stanford.edu/comptop/references/dw.pdf

[3] M. Titsias and N. Lawrence, "Bayesian gaussian process latent variable model," *Artificial Intelligence*, vol. 9, pp. 844–851, 2010. [Online]. Available: http://eprints.pascal-network.org/archive/00006343/

[4] S. Shanmuganathan, P. Sallis, and J. Buckeridge, "Self-organising map methods in integrated modelling of environmental and economic systems," *Environmental Modelling Software*, vol. 21, no. 9, pp. 1247–1256, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/B6VHC-4H2G0BT-1/2/c6a76a6e0c49e363de2db3c0bd83f84c

[5] G. Liao, W. Li, T. Shi, and R. B. K. N. Rao, "Application of generative topographic mapping to gear failures monitoring," *International Journal of COMADEM*, vol. 5, no. 3, 2002.

[6] T. Kohonen, *Self-Organising Maps*, 3rd ed. Berlin: Springer-Verlag, 2001.

[7] E. Lopez-Rubio, "Probabilistic self-organizing maps for continuous data." *IEEE Transactions on Neural Networks*, vol. 21, no. 10, pp. 1543–1554, 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20729166

[8] N. V. Swindale, "The development of topography in the visual cortex: a review of models," *Network Computation in Neural Systems*, vol. 7, no. 2, pp. 161–247, 1996. [Online]. Available: http://dx.doi.org/10.1088/0954-898X/7/2/002

[9] G. J. Goodhill, "Contributions of theoretical modeling to the understanding of neural map development." *Neuron*, vol. 56, no. 2, pp. 301–311, 2007. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17964247

[10] E. Erwin, K. Obermayer, and K. Schulten, "Models of orientation and ocular dominance columns in the visual cortex: A critical comparison," *Neural Computation*, vol. 7, no. 3, pp. 425–468, 1995. [Online]. Available: http://www.mitpressjournals.org/doi/pdf/10.1162/neco.1995.7.3.425

[11] G. Mitchison, "A type of duality between self-organizing maps and minimal wiring," *Neural Computation*, vol. 7, no. 1, pp. 25–35, 1995. [Online]. Available: http://www.mitpressjournals.org/doi/abs/10.1162/neco.1995.7.1.25

[12] D. B. Chklovskii and A. A. Koulakov, "Maps in the brain: what can we learn from them?" *Annual Review of Neuroscience*, vol. 27, no. Mitchison 1991, pp. 369–392, 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15217337

[13] R. Durbin and G. Mitchison, "A dimension reduction framework for understanding cortical maps," *Nature*, vol. 343, no. 6259, pp. 644–647, 1990. [Online]. Available: http://connes.berkeley.edu/ amir/vs298/durbin-mitchison.pdf

[14] C. E. Giacomantonio and G. J. Goodhill, "The effect of angioscotomas on map structure in primary visual cortex." *Journal of Neuroscience*, vol. 27, no. 18, pp. 4935–4946, 2007. [Online]. Available: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1791-10.2011

[15] M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust, "Do we know what the early visual system does?" *Journal of Neuroscience*, vol. 25, no. 46, pp. 10 577–97, 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16291931

[16] D. C. Knill and A. Pouget, "The bayesian brain: the role of uncertainty in neural coding and computation." *Trends in Neurosciences*, vol. 27, no. 12, pp. 712–719, 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15541511

[17] W. J. Ma, J. M. Beck, P. E. Latham, and A. Pouget, "Bayesian inference with probabilistic population codes." *Nature Neuroscience*, vol. 9, no. 11, pp. 1432–1438, 2006. [Online]. Available: http://discovery.ucl.ac.uk/7609/

[18] M. W. Spratling, "Predictive coding as a model of response properties in cortical area v1." *Journal of Neuroscience*, vol. 30, no. 9, pp. 3531–3543, 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20203213

[19] K. Friston, "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20068583

[20] K. Friston, J. Kilner, and L. Harrison, "A free energy principle for the brain," *Journal of PhysiologyParis*, vol. 100, no. 1-3, pp. 70–87, 2006. [Online]. Available: http://discovery.ucl.ac.uk/128290/

[21] K. Friston and K. Stephan, "Free-energy and the brain," *Synthese*, vol. 159, no. 3, pp. 417–458, 2007. [Online]. Available: http://discovery.ucl.ac.uk/143035/

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood for incomplete data via the em algorithm," *Journal of the Royal Statistical Society Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[23] M. Carreira-Perpinan, P. Dayan, and G. Goodhill, "Differential priors for elastic nets," *Proc of the 6th Int Conf Intelligent Data Engineering and Automated Learning IDEAL05*, pp. 335–42, 2005. [Online]. Available: http://discovery.ucl.ac.uk/97508/

[24] M. Carreira-Perpinan and G. J. Goodhill, "Generalised elastic nets," *Convergence*, no. section 6, p. 52, 2011. [Online]. Available: http://arxiv.org/abs/1108.2840

[25] C. M. Bishop, M. Svensen, and C. K. I. Williams, "Developments of the generative topographic mapping," *Neurocomputing*, vol. 21, no. 1-3, pp. 203–224, 1998. [Online]. Available: http://eprints.aston.ac.uk/1232/

[26] J. Q. Jong Youl Choi, M. E. Pierce, and G. Fox, "Generative topographic mapping by deterministic annealing," *Int. Conf. on Membrane Computing*, pp. 47–56, 2010.

[27] I. Olier and A. Vellido, "Variational bayesian generative topographic mapping," *Journal of Mathematical Modelling and Algorithms*, vol. 7, no. April, pp. 371–387, 2008.

[28] I. Olier, A. Vellido, and J. Giraldo, "Kernel generative topographic mapping," *In Procs. of the 18th European Symposiun on Artificial Neural Networks (ESANN 2010)*, pp. 481–486, 2010.

[29] C. M. Bishop, *Pattern Recognition and Machine Learning*, M. Jordan, J. Kleinberg, and B. Schlkopf, Eds. Springer, 2006, vol. 4, no. 4. [Online]. Available: http://www.library.wisc.edu/selectedtocs/bg0137.pdf

[30] MATLAB, *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.

[31] M. Carreira-Perpinan. (2012, Jan) Published papers by miguel carreira-perpinan. [Online]. Available: http://faculty.ucmerced.edu/mcarreira-perpinan/papers.html

[32] I. T. Nabney and C. M. Bishop, "Netlab neural network software," *Synapse*, pp. 15–18, 1997. [Online]. Available: http://www.ncrg.aston.ac.uk/netlab/

[33] G. Goodhill and D. Willshaw, "Application of the elastic net algorithm to the formation of ocular dominance stripes," *Network*, vol. 1, no. 1, pp. 41–59, 1990.

[34] D. Cornford, M. Schroeder, and I. T. Nabney, "Data visualisation and exploration with prior knowledge," *Engineering Applications of Neural Networks 11th International Conference EANN 2009 London UK August 2729 2009 Proceedings*, p. 131, 2009. [Online]. Available: http://www.springerlink.com/content/p63133k8561714q4/