J Neurophysiol 92: 2947–2959, 2004;
10.1152/jn.00281.2004.

Influence of Lateral Connections on the Structure of Cortical Maps

Miguel Á. Carreira-Perpiñán and Geoffrey J. Goodhill
Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007

Submitted 22 March 2004; accepted in final form 2 June 2004

Carreira-Perpiñán, Miguel Á. and Geoffrey J. Goodhill. Influence of lateral connections on the structure of cortical maps. *J Neurophysiol* 92: 2947–2959, 2004; 10.1152/jn.00281.2004. Maps of ocular dominance and orientation in primary visual cortex have a highly characteristic structure. The factors that determine this structure are still largely unknown. In particular, it is unclear how short-range excitatory and inhibitory connections between nearby neurons influence structure both within and between maps. Using a generalized version of a well-known computational model of visual cortical map development, we show that the number of excitatory and inhibitory oscillations in this interaction function critically influences map structure. Specifically, we demonstrate that functions that oscillate more than once do not produce maps closely resembling those seen biologically. This strongly suggests that local lateral connections in visual cortex oscillate only once and have the form of a Mexican hat.

INTRODUCTION

The primary visual cortex of cats, monkeys, and ferrets is characterized by a series of columnar maps, each representing particular features of the visual input. The most noticeable and well-characterized of these are visual field position (VF, topography), ocular dominance (OD), and orientation preference (OR). These maps coexist on the same neural substrate in highly stereotyped patterns. For instance, OD columns tend to intersect OR columns at steep angles, and OR pinwheels tend to lie at the center of OD columns (Bartfeld and Grinvald 1992; Crair et al. 1997; Hübener et al. 1997; Kim et al. 1999; Müller et al. 2000; Obermayer and Blasdel 1993). These maps and their interrelationships have provided a paradigm example for the investigation of how cortical structure arises during development.

In particular, theoretical models of columnar development have been remarkably successful at accounting for many of the observed features of cortical maps (reviewed in Carreira-Perpiñán and Goodhill 2002b; Erwin et al. 1995; Swindale 1996). Many of these models can be seen, explicitly or implicitly, as trading off coverage versus continuity (Carreira-Perpiñán and Goodhill 2002a; Swindale 1996; Swindale et al. 2000). *Coverage* refers to a desire to represent all input features uniformly, whereas *continuity* refers to the desire to represent all input features smoothly. In the cortex, continuity arises from the pattern of short-range lateral connections. By specifying the degree to which neurons at different distances from a particular cortical neuron are excited or inhibited by that neuron, the continuity term in theoretical models effectively controls the type of surface defined by the sheet of cortical receptive fields through the higher-dimensional space of input features, and thus the relations between maps.

A large number of experimental studies have investigated the form these lateral connections take in primary visual cortex (V1; reviewed in Callaway 1998; Lund et al. 2003). Anatomically, a localized injection of an axonal tracer such as biocytin (e.g., Amir et al. 1993; Bosking et al. 1997; Kisvárday et al. 1997) or GFP (green fluorescent protein) adenovirus (Stettler et al. 2002) leads to a small, densely labeled central region of diameter roughly 500 μm, surrounded by a halo of more patchy and diffuse connectivity. Physiological studies investigating the sign of the connections within the central region have generally found a mixture of excitation and inhibition (e.g., Hata et al. 1991; Roerig and Chen 2002). It has also been suggested that within this region excitation is more spatially restricted than inhibition (Hata et al. 1991), although this is not generally confirmed (see also Swindale 1996). Whereas the short-range connections are largely nonspecific, the longer-range connections tend to link cells with similar orientation preferences (e.g., Bosking et al. 1997; Roerig and Chen 2002). Although the anatomical data do not directly support a lateral-connection pattern with a Mexican-hat form (central excitation surrounded by inhibition), other sources of evidence exist. For example, in mouse visual cortex, extracellular recordings and optical imaging (see Schuett et al. 2002 and references therein) have shown that stimulation of a limited region of the visual cortex also elicits surround inhibition, thus suggesting a Mexican-hat pattern. The range of inhibition also seems to extend farther than suggested by the purely anatomical evidence, and this has a role in map development (see Hensch and Stryker 2004; Fagiolini et al. 2004 and references therein).

Models of cortical pattern formation during development have generally considered only the effect of the dense, short-range mixture of excitatory and inhibitory connections, and have ignored the patchy longer-range connections. In particular, most models have assumed that lateral interactions take the form of a radially symmetric disk of short-range excitation surrounded by an annulus of longer-range inhibition (e.g., Goodhill 1993; Miller et al. 1989; von der Malsburg 1973; Willshaw and von der Malsburg 1976; reviewed in Swindale 1996). Even models that allow plasticity of lateral interactions have imposed the condition of short-range excitation and longer-range inhibition (Sirosh and Miikkulainen 1997). Although as discussed above there is little direct biological evidence to support this assumption, such "Mexican-hat" functions for short-range connectivity are very popular in models for cortical function as well as cortical development (e.g.,

Address for reprint requests and other correspondence: G. J. Goodhill, Dept. of Neuroscience, Georgetown University Medical Center, 3900 Reservoir Road NW, Washington, DC 20007 (E-mail: geoff@georgetown.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Dragoi and Sur 2000; Ernst et al. 2001; Kang et al. 2003; Somers et al. 1998). One very appealing feature of Mexican-hat interactions is their powerful pattern-forming properties (Ball 1999; Meinhardt 1982; Turing 1952).

However, despite this popularity there has been very little investigation of how important the precise form of short-range Mexican-hat interactions (i.e., the continuity term) is to theoretical models of cortical development. If it were the case that different types of continuity term led to maps with broadly similar properties, then models would not provide strong constraints on what patterns of lateral connections might exist in the real cortex. Here we show that, on the contrary, the precise form of the continuity term provides very sensitive control over the details of the maps formed. This greatly strengthens the prediction that lateral connections in the real cortex do indeed have the form of short-range excitation and longer-range inhibition. Our results demonstrate the fragility of the interrelations between maps with respect to a class of Mexican-hat functions that could be implemented in biological hardware in terms of localized arbors (of different widths and amplitudes) of excitatory and inhibitory connections.

The particular model we investigate is the elastic net (Dayan 1993; Durbin and Mitchison 1990; Durbin and Willshaw 1987; Durbin et al. 1989; Goodhill and Willshaw 1990). Together with the Kohonen model, the elastic net exceeds all other models in the closeness with which the maps it produces match the details of real maps (Erwin et al. 1995; Swindale 1996). Unlike the Kohonen model, however, the elastic net has an objective function that explicitly implements the trade-off between coverage and continuity. This property allows the effect of the continuity term to be cleanly separated from that of the coverage term. We first present a novel interpretation of the continuity term in terms of a pattern of lateral interactions in the cortex and show that the standard form of the continuity term, in effect a first-order derivative, represents short-range excitation and longer-range inhibition. We also generalize the continuity term to higher-order derivatives and show how these correspond to patterns of lateral interactions that oscillate more than once (i.e., the annulus of inhibition is now surrounded at longer range by an annulus of excitation, and so forth). Simulations of the joint development of OD and OR columns are then compared using these differing assumptions regarding lateral interactions. We find that as derivative order increases, normal map structure is gradually disrupted: in particular many OR pinwheels move to OD borders, and the histogram of intersection angles between OR and OD columns becomes flat, although the maps themselves remain columnar. These results lead to two testable biological predictions: *1*) the sign of short-range lateral interactions changes only once in visual cortices that have been well-characterized so far; *2*) if cortices are found where columns do not intersect at right angles and OR pinwheels are not constrained to mostly lie on OD centers, these cortices may contain short-range lateral interactions that change sign more than once.

METHODS

The elastic net (Durbin and Mitchison 1990; Durbin and Willshaw 1987; Goodhill and Willshaw 1990) produces maps that attempt to minimize a trade-off between coverage C, the desire to represent all stimulus features uniformly in the cortex (Swindale 1991), and continuity R, the desire to produce a smooth mapping of stimulus features. This trade-off is expressed in terms of minimizing an "energy function" $E = C + (\beta/2)R$, where β controls the weight of R with respect to C. Let $\mathbf{x}_n$ represent the stimulus space vector n (combination of preferred position, ocular dominance, and orientation), and $\mathbf{y}_m$ represent the center in stimulus space of the receptive field of cortical neuron m. [Note: Although it is convenient to refer to m as being a single cortical neuron, it should more properly be thought of as representing a small patch of neurons. We use the term "receptive field" to refer to all features jointly (i.e., visual field, ocular dominance, and orientation), and not only to the visual field.] In the elastic net the coverage term is then defined as follows

$$C(\mathbf{y}_1, \ldots, \mathbf{y}_M; K) = -K \sum_{n=1}^{N} \log \sum_{m=1}^{M} \exp\left(-\frac{1}{2}\left\|\frac{\mathbf{x}_n - \mathbf{y}_m}{K}\right\|\right)^2$$

where K is a parameter that gives the receptive field size in feature space ($\|\ldots\|$ represents the Euclidean norm, or length, of a vector), such that large values of K mean cortical neurons are quite unselective for specific features, whereas small values of K mean they are quite selective. We reduce (anneal) K over the course of the simulation, corresponding to a gradual sharpening of the cortical receptive fields. This implements an efficient optimization procedure known as *deterministic annealing* (Rose 1998). This iterative minimization procedure can be interpreted in terms of Hebbian learning (Durbin and Mitchison 1990), although other interpretations are also possible. The stimulus space is densely populated with points along the dimensions of visual field (VFx, VFy), ocular dominance (OD), and orientation preference and selectivity (OR, ORr). In our case, the stimulus values (training set) are a grid of $N_x \times N_y \times N_{OD} \times N_{OR} \times 1$ in the rectangle $[0, 1] \times [0, 1] \times [-l, l] \times [-(\pi/2), (\pi/2)] \times [0, r]$, where OR is conventionally represented by 2 variables in polar coordinates, with N_{OR} values uniformly arranged on a ring of radius r (see e.g., Swindale 1996). There are $N = N_x N_y N_{OD} N_{OR}$ such 5-dimensional points $\mathbf{x}_n$. The net consists of a square lattice with M centroids, representing a square array of cortical neurons (the depth dimension of the cortex is not modeled).

The continuity term in the elastic net is defined as

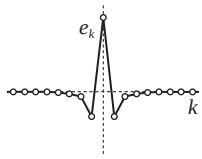
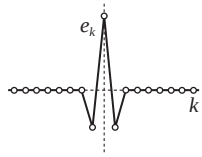
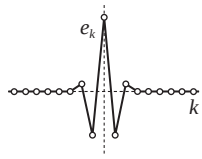
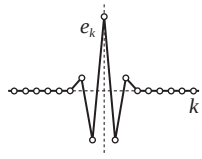
$$R(\mathbf{y}_1, \ldots, \mathbf{y}_M) = \sum_{m=1}^{M} \|\mathbf{f}(\mathbf{y}_m)\|^2$$

where for each neuron m the term $\mathbf{f}$ is a fixed linear combination of cortical neurons in a neighborhood of m. That is, $\mathbf{f}$ specifies the strength of the interaction between a cortical neuron and its neighbors. We call the coefficients of the linear combination the *stencil*. In the original formulation of the elastic net (Durbin and Willshaw 1987), R was defined as $\sum_{m=1}^{M} \|\mathbf{y}_{m+1} - \mathbf{y}_m\|^2$, that is, the sum of squared distances in feature space between adjacent receptive field centers of neighboring cortical neurons. In the generalized elastic net (Carreira-Perpiñán and Goodhill, unpublished observations), each term $\mathbf{f}(\mathbf{y}_m) = (\mathbf{y}_{m+1} - \mathbf{y}_m)$ is viewed as a discretized derivative operator of order $p = 1$ (i.e., an approximate gradient). We can prove (Carreira-Perpiñán and Goodhill, unpublished observations) that the continuity term can be rewritten as a Mexican-hat function as follows (see $p = 1$ in Table 1). Define the *cortical interaction function* e_k and the local linear combination $\mathbf{g}$ with coefficients e_k as

$$e_k = \frac{-4}{\pi(4k^2 - 1)} \quad k = 0, \pm 1, \pm 2, \ldots \quad \mathbf{g}(\mathbf{y}_m) = \sum_{k=-\infty}^{\infty} e_k \mathbf{y}_{m+k} \qquad (1)$$

Then, for a large number M of net centroids, it holds that

TABLE 1. *Cortical interaction functions for continuity terms of differential order $p = 1$ to 4 in the generalized elastic net in 1D*

Order p	Continuity Term $R = \sum_{m=1}^{M} \|\mathbf{f}(\mathbf{y}_m)\|^2$	Equivalent Mexican-Hat Graph	Cortical Interaction Function e_k
1	$\sum_m \|\mathbf{y}_{m+1} - \mathbf{y}_m\|^2$		$e_k = \dfrac{-4}{\pi(4k^2-1)}$, $k = 0, \pm 1, \pm 2, \cdots$
2	$\sum_m \|\mathbf{y}_{m+1} - 2\mathbf{y}_m + \mathbf{y}_{m-1}\|^2$		$e_k = \begin{cases} 2, & k = 0 \\ -1, & k = \pm 1 \\ 0, & \text{rest} \end{cases}$
3	$\sum_m \|\mathbf{y}_{m+2} - 3\mathbf{y}_{m+1} + 3\mathbf{y}_m - \mathbf{y}_{m-1}\|^2$		$e_k = \dfrac{96}{\pi(4k^2-1)(4k^2-9)}$, $k = 0, \pm 1, \pm 2, \cdots$
4	$\sum_m \|\mathbf{y}_{m+2} - 4\mathbf{y}_{m+1} + 6\mathbf{y}_m - 4\mathbf{y}_{m-1} + \mathbf{y}_{m-2}\|^2$		$e_k = \begin{cases} 6, & k = 0 \\ -4, & k = \pm 1, \\ 1, & k = \pm 2 \\ 0, & \text{rest} \end{cases}$

The equations for R constitute the discretized difference formulation (using the **f** linear combinations), whereas the graphs show the equivalent Mexican-hat representation and corresponding cortical interaction function (using the **g** linear combinations). The vertical axes have been rescaled to aid the visualization. All functions are constrained to integrate to 0 (note the sum of the coefficients of the discretized difference is 0 in each row). For $p = 1$, the cortical interaction function has only one change of sign (considering only one half of the horizontal axis, by symmetry); as p increases, the function develops progressively more changes of sign (oscillations). For a 2D net, the resulting lateral connections consist of several concentric regions of alternating excitation and inhibition.

$$R(\mathbf{y}_1, \ldots, \mathbf{y}_M) = \sum_{m=1}^{M} \|\mathbf{y}_{m+1} - \mathbf{y}_m\|^2 \approx \sum_{m=1}^{M} \|\mathbf{g}(\mathbf{y}_m)\|^2$$

(Exact expressions can be derived for any finite M, but are complicated and nearly coincide numerically with the limit approximation.) In other words, we can represent the continuity term $R(\mathbf{y}_1, \ldots, \mathbf{y}_M)$ either in terms of the original linear combinations $\mathbf{f}(\mathbf{y}_m) = \mathbf{y}_{m+1} - \mathbf{y}_m$ or in terms of the new linear combinations $\mathbf{g}(\mathbf{y}_m)$ given in *Eq. 1*. That is, the continuity term in the original elastic net is mathematically equivalent to a Mexican-hat function consisting of short-range excitation and long-range inhibition. This demonstrates the equivalence in the elastic net model of the Mexican-hat cortical interaction and the interpretation of Durbin and Mitchison (1990) in terms of wire-length minimization (in the stimulus space). It also shows that the sum-of-squared-distances equation, which involves only nearest-neighbor connections, can be implemented with a denser net of connections (whose strength decays rapidly with distance).

Instead of choosing, for the function $\mathbf{f}(\mathbf{y}_m)$, the first-order derivative of sensory space over cortex, we could explore the effects of choosing higher-order derivatives. For this paper we chose derivatives of orders 2, 3, and 4. The discrete forms of these higher-order derivatives are shown in Table 1; they are obtained by repeatedly applying the first-order derivative $(\mathbf{y}_{m+1} - \mathbf{y}_m)$. In each case, we can show that the sum $\sum_{m=1}^{M} \|\mathbf{f}(\mathbf{y}_m)\|^2$ is equivalent, in the limit of large M, to $\sum_{m=1}^{M} \|\mathbf{g}(\mathbf{y}_m)\|^2$ with an appropriate **g** interaction function; and that this **g** interaction function has a progressively larger number of distal oscillations, or alternating rings of excitatory and inhibitory connections, with increasing order of the derivative. Because the continuity term R should not depend on the location of the net, but only the relative arrangement of its centroids, the coefficients in the stencil should add to zero, which is the same as requiring that applying the stencil to a constant function give zero (by definition of differential stencil). 2D stencils can be readily obtained by passing a 1D stencil horizontally and vertically, resulting in 2 terms, R_h and R_v, that are then added. This is the natural extension of the first-order 2D stencil used in the original elastic net, although it leads to a Fourier spectrum that is increasingly anisotropic as p increases. There are other ways to construct 2D stencils (e.g., the Laplacian stencil)

$$\nabla^2_+ = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

of order $p = 2$, which has a more isotropic spectrum. Note however that in a discrete grid there is not a unique definition of the concept of isotropy. Although in reality it is likely that the lateral connections are

maturing as the maps develop (Sur and Leamey 2001), the elastic net model (in common with most other models) simplifies this situation by assuming a fixed pattern of lateral connections.

Simulation setup

We trained nets for derivative orders $p = 1$, 2, 3, and 4 with the following configuration: training set: $N_x = N_y = 20$, $N_{OD} = 2$, $N_{OR} = 12$, $l = 0.09$, and $r = 0.16$ (totaling $N = 9{,}600$ points in 5 dimensions); net: 128×128 centroids; $\beta = 10$; nonperiodic boundary conditions. We reduced K with an annealing rate of 0.992 from 0.1 to the point at which the OD and OR maps have just arisen ($K \approx 0.05$ for all values of p). We achieved the minimization of E at each value of K by Cholesky factorization (Carreira-Perpiñán and Goodhill, unpublished observations). This method is both more efficient and more robust than the numerical procedures used previously (e.g., Durbin and Mitchison 1990; Durbin and Willshaw 1987; Goodhill and Willshaw 1990), which were limited to very small β values. We ran 5 different simulations for each value of p to obtain error bars in Figs. 3 and 4, adding a small amount of random noise to the initial net and the training set in each case. The simulations were performed using custom software written in Matlab.

Several of the statistics reported are based on histograms; in all cases we used $n = 10$ bins and took each bin's center as class representative. Using a different number of bins (100 instead of 10) did not significantly alter any of the results. To quantitatively characterize the histograms (average and variability) as a function of p in a compact way in Fig. 3, we summarized each histogram by 3 statistics: *1*) its Kullback–Leibler (KL) divergence (Cover and Thomas 1991) to the uniform histogram in the same domain, defined as $\mathrm{KL}(u) = \sum_{i=1}^{n} p_i \log (p_i/u_i)$, where p_i and u_i are the values (normalized to sum 1) at bin i of the histogram and the uniform histogram, respectively; $\mathrm{KL}(u)$ quantifies departure from uniformity and takes values from 0 for a uniform histogram to $\log n \approx 2.303$ when all p_i are zero except at a single bin; *2*) its Fisher skewness $\gamma_1 = \mu_3 \mu_2^{-3/2}$ (where μ_i is the ith central moment), which is 0 for a histogram symmetric around its mean and negative (positive) for tails on the left (right) of the mean; *3*) its mean μ. Both γ_1 and μ are independent of the number of bins n; all 3 are independent of translations or scalings of the data. Other summary statistics are possible, but these 3 sufficed to characterize the histograms we obtained.

Stripe width

The power of the discrete Fourier transform of OD and OR angle maps was roughly isotropic and concentrated around a ring. We summarized it by the mean wavelength (mean stripe width), averaged over all directions.

Crossing angles

We disregarded the points lying 5 pixels or less from the boundaries of the net (a thin inner stripe framing the maps in Fig. 1) to eliminate border effects. At each pixel, we computed the angle between the gradient vectors of OD and OR (see following text) and mapped it to [0°, 90°], thus obtaining the angle between contours of OD and OR. Each such angle counted with a weight proportional to the product of the gradient moduli of OD and OR, $\| \nabla \mathrm{OD} \| \times \| \nabla \mathrm{OR} \|$, so that pixels lying in areas of either constant OD or constant OR (i.e., away from borders of OD or contours of OR) were effectively removed from the histogram. This is because, in an area where either map is nearly constant, the gradient vector is negligible in modulus and its direction basically arbitrary, unlike along borders, where the gradient vector is large and its direction well defined.

Pinwheel distributions

An OR singularity, or pinwheel, is defined positive if the orientation angle increases in a clockwise direction around the pinwheel and negative if anticlockwise. Pinwheels were automatically located in the OR maps as follows. First, the winding number of each pixel in the OR map was estimated by summing the increments of OR angle (in $[-(\pi/2), (\pi/2)]$) along a closed path (a square of radius 1 pixel centered in the pixel considered) in a clockwise direction and dividing the result by 2π; this results in 0 for nonpinwheel points and $+½$ ($-½$) for pixels at or adjacent to a positive (negative) pinwheel, respectively. The exact pinwheel location was obtained by grouping clusters of nonzero winding number and computing their centers of mass. To quantify the layout of pinwheels with respect to OD stripes, we computed, for every pinwheel, the distribution of distances of a given pinwheel to its closest OD border. We prefer this to calculating distances to OD centers because the "center of an OD stripe" is generally not well defined, except for stripes that are translations of each other, and becomes difficult to use with stripes with irregular boundaries, forks, islands, or other complex shapes. In contrast, the OD borders are well defined in all these cases.

To compute the distances between a pinwheel and its closest OD border, we represented the OD borders by a finely spaced collection of points (the contours generated by Matlab for the OD map); the said distance was then given by the point in this collection closest to the given pinwheel. We considered that a pinwheel lay on an OD border if the distance between the 2 was 1 pixel or less. The distributions of random pinwheels were built by generating, uniformly over the cortex, the same number of pinwheels as in the actual map produced by the simulation, and then computing the distance histogram; this process was repeated 50 times to obtain an average histogram with error bars that is the one plotted along with the actual ones for comparison. Note that theoretical distributions (such as the Rayleigh distribution for nearest-neighbor pinwheels; Müller et al. 2000) differ from the true one because they do not take into account the border effects of a finite cortex.

Gradient correlations

The gradient of a variable f (one of VFx, VFy, OD, OR) with respect to the cortical location (x, y)

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

was approximated at each pixel (i, j) by a central finite difference

$$\nabla f_{ij} = \left(\frac{f_{i+1,\,j} - f_{i-1,\,j}}{2h}, \quad \frac{f_{i,j+1} - f_{i,\,j-1}}{2h} \right)$$

where $h = 1$ pixel is the grid size. For points along the cortex boundaries, a forward difference

$$\frac{f_{i+1,\,j} - f_{i,j}}{h}$$

was used at a small accuracy loss. In both cases, for OR the gradient values were corrected to have a maximal absolute value of $(\pi/2)$. The 2D gradient vector ∇f was converted to polar coordinates (modulus, angle) for each variable separately, with the angle being used in the histogram of crossing angles. The moduli were used for the correlation plots (the modulus of VF gradient was defined operationally as the sum of the gradient moduli of VFx and VFy). The units of the gradient modulus of variable f are those of variable f per pixel. The linear correlation was quantified by Pearson's correlation coefficient r. The customary regression lines $X \rightarrow Y$ and $Y \rightarrow X$ minimize the vertical and horizontal errors, whereas the line X, Y minimizes the orthogonal

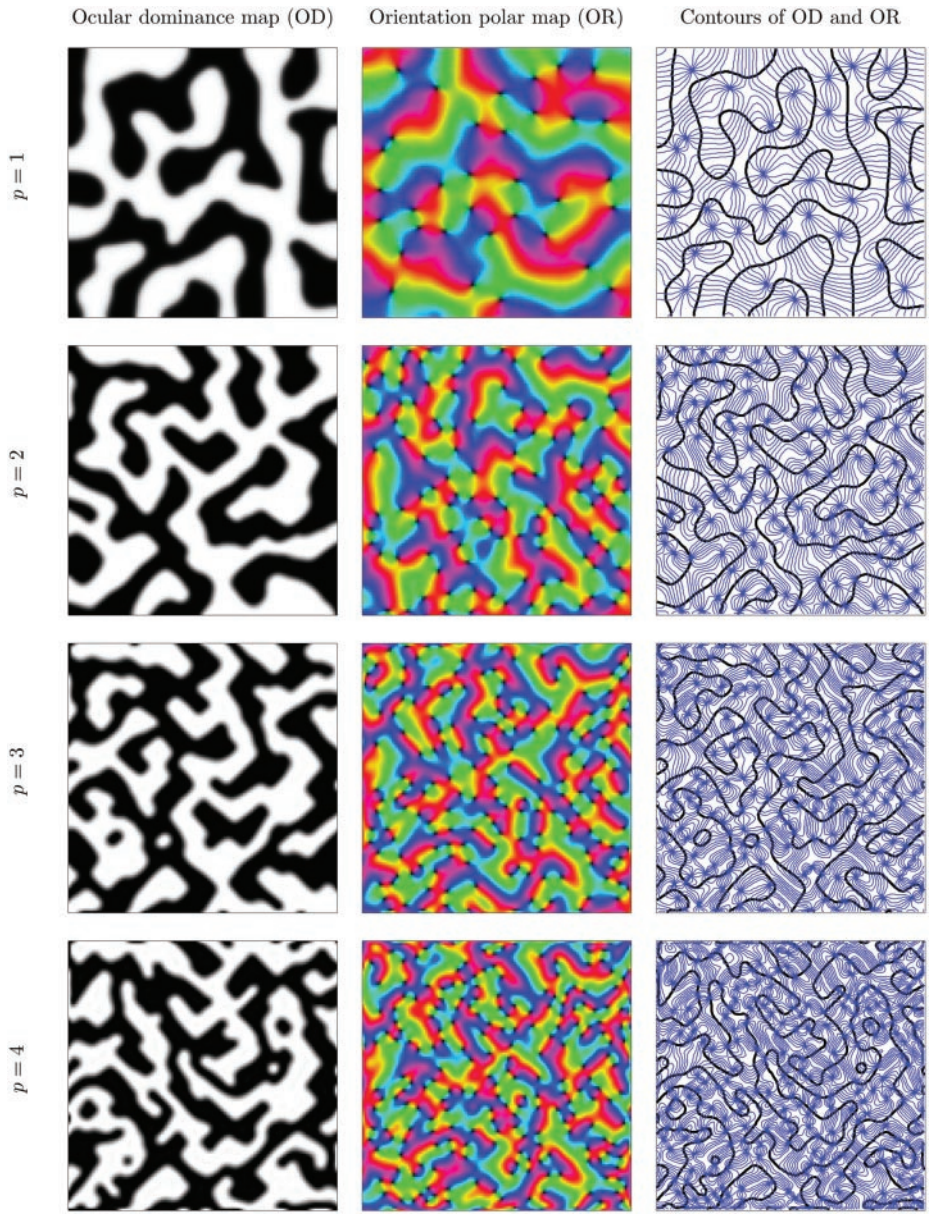


FIG. 1. Simulated maps of ocular dominance (OD) and orientation preference (OR) for a generalized elastic net of derivative orders $p = 1$ to 4. *Left column*: OD map. *Center column*: orientation polar map. *Right column*: contours of ocular dominance and orientation. For each order p, the continuity terms and cortical interaction functions are given in Table 1. Grids contain 128×128 cortical points. This is one of 5 simulations run with different noise in the initial net and training set. Training set consists of a uniform, dense grid in the 5-dimensional stimulus space of visual field (VFx, VFy), OD and orientation (selectivity ORr and preferred angle ORt). Maps are the result of minimizing the generalized elastic net energy $E = C + (\beta/2)R$ with $\beta = 10$. Note how as p increases the pinwheels increasingly co-locate with OD borders, and the contours increasingly align along the diagonals. For $p = 1$, the pinwheels are preferentially away from OD borders, near OD centers, and the contours of OD and OR intersect preferentially at right angles. As p increases, an increasing number of pinwheels lie right on OD borders, like beads on a string, and the contours tend to intersect at any angle.

distances to the line and is given by the principal component of the data covariance matrix. All 3 lines pass through the data mean but with different slopes in general. For highly correlated data ($r^2 \approx 1$), all 3 lines are very close, whereas for highly uncorrelated data ($r^2 \approx 0$) $X \rightarrow Y$ and $Y \rightarrow X$ become horizontal and vertical, respectively.

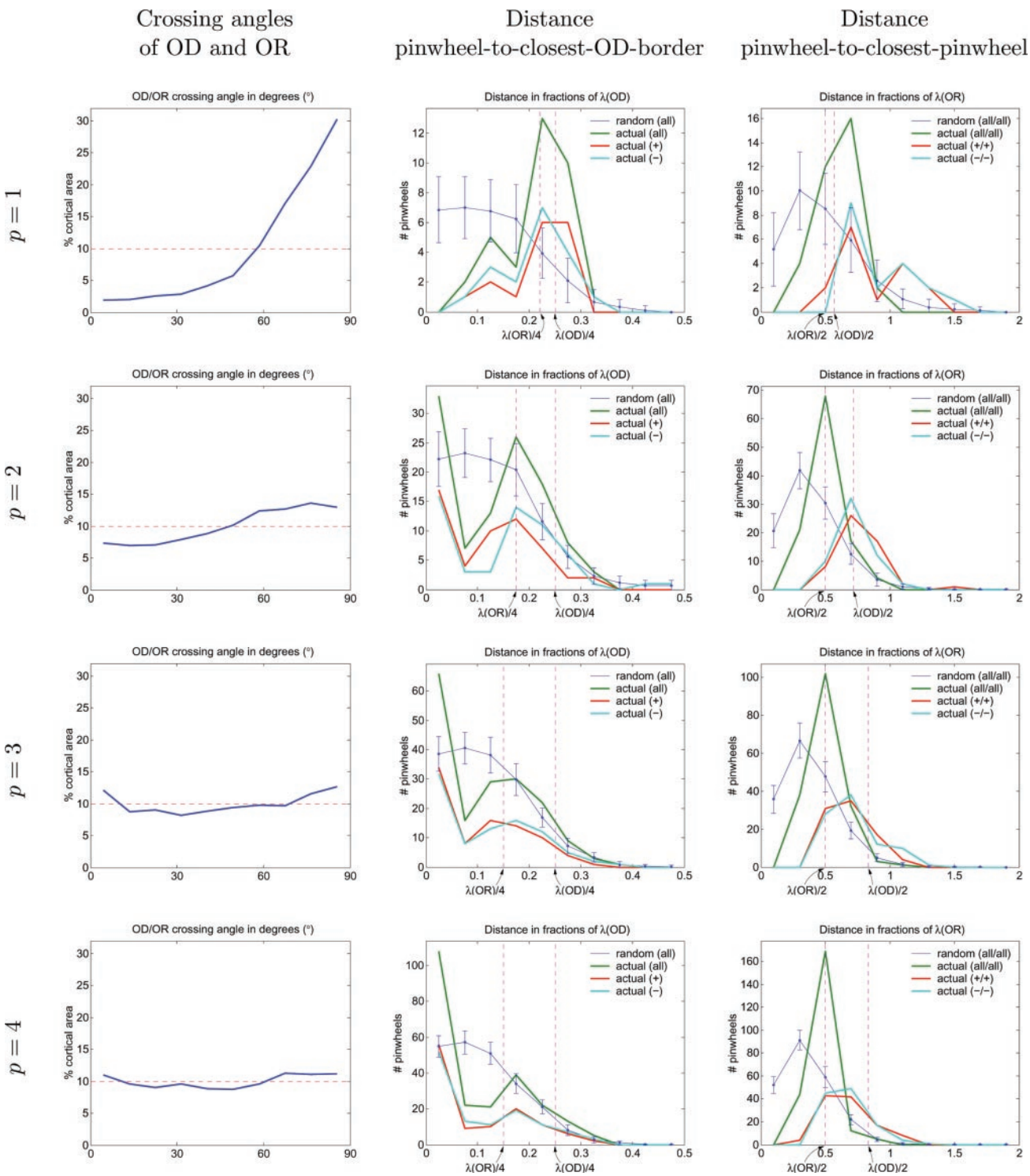
Gradient correlations have also been reported in the experimental literature (e.g., Das and Gilbert 1997; Hetherington and Swindale 1999), where the preferences of VF and OR were determined electrophysiologically at several cortical sites and then the difference in preference between neighboring sites divided by their cortical distance. However, for practical reasons "neighboring" sites are much further apart over the cortical surface in these studies than neighboring points in our simulated cortex.

RESULTS

We present the results obtained with a generalized elastic net model of the maps of visual field position (VF), ocular dominance (OD), and orientation preference (OR). The model attempts to represent the 5 dimensions of VF (x and y directions), OD and OR (preferred angle and selectivity) using a 2D array of cortical neurons (the elastic net), thus achieving a reduction of dimension. The maps are the result of maximizing the coverage of these stimuli while minimizing a continuity constraint on the shape of net (the geometry of the maps). The continuity constraint penalizes the squared magnitude of the derivative of order p integrated over the whole net, thus forcing a representation of the maps that is as smooth as possible. Biologically, the continuity constraint represents a cortical interaction function with the shape of a Mexican hat, in which the number of excitatory–inhibitory regions increases with the derivative order p (Table 1). We considered $p = 1$ to 4. Table 1 shows, for each value of p from 1 to 4, the corresponding form of the continuity term in the elastic net, and the equivalent (short-range) cortical interaction function (for details see METHODS). For $p = 1$ (the standard elastic net) the cortical interaction

function has the form of short-range excitation surrounded by an annulus of longer-range inhibition, which decreases smoothly to zero. For $p = 2$, the magnitude of the longer-range inhibition decreases quickly to zero. For $p = 3$, the magnitude of the longer-range inhibition decreases quickly to zero, and then there is a small annulus of excitation surrounding this, which decreases smoothly to zero. For $p = 4$, the annulus of excitation is stronger but falls rapidly to zero. It should be emphasized that in each case these functions represent only radially symmetric short-range cortical interactions (i.e., within a radius of about 500 microns) and longer-range patchy connections are not considered.

By simulating the development of VF, OD, and OR maps using the elastic net with each of these continuity terms, we demonstrate that these very subtle differences in cortical interactions, in the form of the Mexican hat, have profound effects



on the detailed structure of the maps obtained. Figure 1 shows the maps obtained from one simulation for derivative orders $p = 1$ to 4. The maps obtained in simulations using different random numbers differed in the detailed layout but were qualitatively similar. Figures 2–5 quantitatively assess the specific maps of Fig. 1, whereas Figs. 3 and 4 summarize the various statistics as a function of the derivative order; the average value and error bars (SDs) result from 5 simulations per order.

Crossing angles

The distribution of crossing angles of the contours of OD and OR for the maps of Fig. 1 are shown in the *left column* of Fig. 2. For $p = 1$, this distribution is strongly skewed toward orthogonality, in agreement with experimental data (macaque: Bartfeld and Grinvald 1992; Obermayer and Blasdel 1993; cat: Crair et al. 1997; Hübener et al. 1997; Kim et al. 1999; Müller et al. 2000). However, as p increases, the distribution quickly loses its bias and becomes practically uniform for $p \geq 3$. The loss of orthogonality is apparent by close inspection of the contours in Fig. 1 (*right column*). The transformation of the distribution from being sharply skewed toward orthogonality (mean 65°) to being practically uniform (mean $\approx$ 45°) is clearly shown in Fig. 3 (*first row*).

Stripe widths

Figure 4 (*left*) plots the mean wavelength of OD and OR in pixels, showing that the stripes become narrower as the derivative order increases. The stripe width depends for any derivative order on the variance along the OD and OR dimensions in the training set, and is related to the order in which both maps arise (Goodhill and Cimponeriu 2000). For the results shown in this paper we chose these variance parameters such that OD columns are slightly wider than OR columns, as is the case in macaques (e.g., Obermayer and Blasdel 1993). However, we also investigated the case where OR columns are wider than OD columns, corresponding to the case in cats (e.g., Löwel et al. 1988). No important differences were found of increasing derivative order in the effects we have quantified here (data not shown). Specifically, all histograms and curves in Figs. 2–5 remained basically the same (same increasing or decreasing pattern, similar location of peaks, etc.), and in Fig. 4 (*left*) the 2 curves cross over between $p = 1$ and $p = 2$ because the OR wavelength decreases faster than the OD one.

It is noticeable that the stripes become increasingly more aligned with the diagonals of the cortical region considered as p increases, as seen from Fig. 1. This is attributed to the fact that the 2D stencils (pattern of lateral interactions) we use are increasingly anisotropic in Fourier space as p increases (Carreira-Perpiñán and Goodhill, unpublished observations). To eliminate the possibility that some of the other effects on map structure we report might not appear using more isotropic 2D stencils representing equal derivative order, we also investigated a 2D Laplacian stencil (order $p = 2$), which has a more isotropic spectrum. The matching of pinwheels and OD borders and the uniformity of crossing angles still appeared with this stencil (results not shown), so we conclude that the anisotropy is not responsible for the effects we report.

Orientation pinwheels

Figure 2 (*right column*) shows the distribution of distances between nearest-neighbor pinwheels, normalized by the wavelength of OR. The histograms are very similar for all orders of p. When pinwheels are considered independently of their sign, the histogram peaks quite sharply at half the OR wavelength and has a slight tail on the right toward higher distances, while quickly dropping to 0 for small distances. If the same number of pinwheels are uniformly distributed at random, then the histogram peaks at a smaller distance and has a much longer right tail. When only pinwheels of the same sign are considered, the histogram contains half the number of pinwheels and has the same shape, but peaks at about ¾ the OR wavelength. This is confirmed by the summary statistics of Fig. 3 (*third row*); the large error bars for $p = 1$ are attributed to the very small number of pinwheels present in the map (about 50), which results in a more variable histogram. A slight tendency for pinwheels to be more closely packed as p increases is visible in Fig. 3 (μ graph), consistent with the slight increase of pinwheels per OR module mentioned above.

This reflects the fact that pinwheels are laid out in a noisy gridlike arrangement (of grid length half the OR wavelength), with pinwheels repelling each other, so that same-sign pinwheels alternate in the nodes of the grid. Thus whereas the numbers of pinwheels and stripe widths vary with p, the arrangement of pinwheels relative to each other remains the

FIG. 2. Distribution of crossing angles of OD and OR, pinwheel-to-closest-OD-border distance and pinwheel-to-closest-pinwheel distance for derivative orders $p = 1$ to 4, for the maps of Fig. 1. Number of bins is 10 for all histograms; the value at each bin is plotted against the center of the bin and the resulting points are connected by lines. *Left column*: histogram of crossing angles in [0°, 90°] of the contours of OD and OR. Horizontal dashed line represents the uniform histogram. Note the bias for orthogonality for $p = 1$ and the progressive flattening as p increases. *Center column*: histogram of the distance from a pinwheel to its closest OD border. Axes represent the number of pinwheels as a function of the distance in fractions of the OD wavelength. Two vertical dashed lines mark ¼ of the wavelengths of OR and OD. "Random (all) " means pinwheels distributed uniformly at random in the map (the SD bars resulting from 50 such random sets); "actual (all) " means the actual pinwheels shown in Fig. 1, irrespective of their sign, with "actual (+) " and "actual (−) " meaning the histogram considering only positive or negative pinwheels, respectively. Note that for $p = 1$ there is a single peak [at around λ(OD)/4, the center of an OD stripe]; but for $p > 1$ there is also a second peak (at 0, the OD border) whose height progressively increases with p at the expense of the OD-center peak. In all cases the histogram is very different from that resulting from random pinwheels. *Right column*: histogram of the distance between nearest-neighbor pinwheels. Axes represent the number of pinwheels as a function of the distance in fractions of the OR wavelength. Two vertical dashed lines mark ½ of the wavelengths of OR and OD. "Random (all/all) " means pinwheels distributed uniformly at random in the map (the SD bars resulting from 50 such random sets); "actual (all/all) " means the actual pinwheels shown in Fig. 1, irrespective of their sign, with "actual (+/+) " and "actual (−/−) " meaning the histogram considering only positive or negative pinwheels, respectively. This histogram remains essentially the same for all values of p.

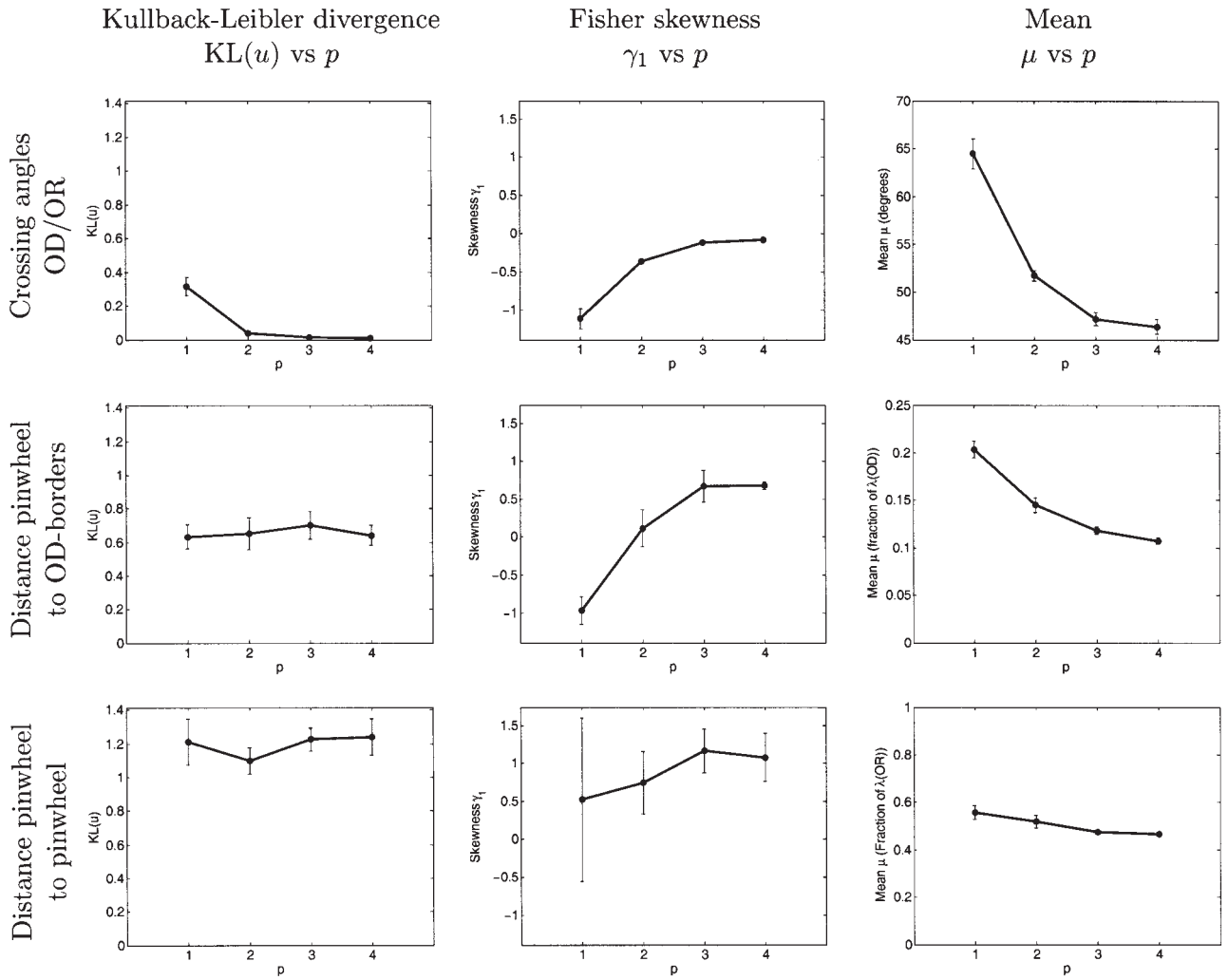


FIG. 3. Summary statistics for the histograms of Fig. 2 as a function of the derivative order $p = 1$ to 4 (explanations of what these statistics mean are in the METHODS). Graphs show the average over 5 different runs of a histogram statistic and the error bars represent SDs. Statistics shown are the Kullback–Leibler (KL) divergence to the uniform histogram KL(u) (*left column*), the Fisher skewness γ_1 (*center column*), and the mean μ (*right column*). Vertical axis for the KL(u) statistic is the same for all rows, as for the γ_1. *Row 1*: statistics for the histogram of crossing angles. All 3 statistics decrease with increasing p, indicating a left-skewed histogram for $p = 1$ that quickly flattens for $p > 1$. *Row 2*: statistics for the histogram of the pinwheel-to-closest-OD-border distance. This histogram remains nonuniform for all p [high KL(u)] but changes the skewness sign and decreases the mean for $p > 1$, indicating the growth of the peak at distance 0 (OD border). *Row 3*: statistics for the histogram of the pinwheel-to-closest-pinwheel distance. This histogram remains essentially unchanged, with a tiny decrease of its mean for increasing p attributed to the increase of the number of pinwheels per OR module.

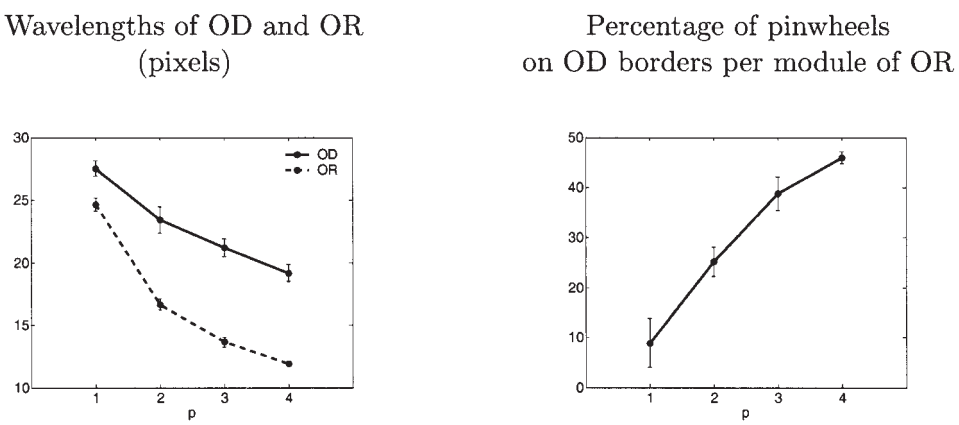


FIG. 4. Summary statistics for map wavelengths and pinwheels on OD borders as a function of the derivative order $p = 1$ to 4. Graphs show the average over 5 different runs of a statistic and the error bars represent SDs. *Left graph*: mean wavelengths of OD and OR in pixels (note the cortical grid is 128×128 pixels). *Right graph*: percentage of pinwheels per module of OR lying right on OD borders, where a module of OR is defined as a square of side equal to the OR wavelength (the plot using the module of OD looked identical). Note how both the OR and OD wavelengths decrease with p and the percentage of pinwheels on OD borders per module of OR increases dramatically from a minority (10%) for $p = 1$ to half for $p = 4$.

same for all p, in agreement with optical imaging data for squirrel and macaque monkeys (Obermayer and Blasdel 1997).

The effect on the distribution of the distance from a pinwheel to its closest OD border is more striking. In Fig. 2 (*center column*), for $p = 1$ the histogram shows a quite sharp peak around ¼ of the OD wavelength, with a long tail on the left. Thus pinwheels tend to locate at the centers of OD columns, almost always avoiding OD borders, as apparent in Fig. 1 and in agreement with optical imaging data (macaque: Bartfeld and Grinvald 1992; Obermayer and Blasdel 1993; cat: Hübener et al. 1997). However, for $p = 2$ the histogram becomes bimodal, preserving a now less-sharp peak at ¼ of the OD wavelength and developing a very sharp peak at distance 0, indicating that pinwheels generally lie either away from the OD borders (at a distance around 0.17 of the OD wavelength) or right on top of an OD border. As p increases, the peak at distance 0, which remains sharp, increases at the expense of the peak at distance 0.17 of the OD wavelength, which becomes broader. This remarkable effect is apparent in Fig. 1: note for $p \geq 2$ how the OD borders meet runs of pinwheels like beads on a string. As Fig. 2 shows, this distribution is very different from that resulting from the same number of pinwheels but uniformly distributed at random, which would have a roughly flat value for distances $\leq$0.15 of the OD wavelength to then taper off at larger distances. The histograms also show that, for all orders, the distance to OD borders is equally distributed with respect to the sign of the pinwheels (naturally, the histograms of positive and negative pinwheels are half the height of the histogram for all signs). The summary statistics in Fig. 3 (*second row*) confirm that the histogram remains nonuniform for all p, its mean μ (normalized by the OD wavelength) monotonically decreases, and its skewness changes from negative for $p = 1$ (tail on the left) to positive for high p (tail on the right). Figure 4 (*right*) shows how the proportion of pinwheels that lie right on OD borders grows from barely 10% for $p = 1$ to almost 50% for $p = 4$. Note that the flattening of the angle histogram is consistent with the increasing proportion of pinwheels on OD borders, given that an OD border crossing a pinwheel must be parallel to some OR contours. In other words, an orthogonality bias requires that pinwheels avoid OD borders.

Gradient correlations

Figure 5 shows the pairwise correlation plots between the gradient moduli of VF, OD, and OR (cf. Bosking et al. 2002; Buzás et al. 2003; Das and Gilbert 1997; Hetherington and Swindale 1999; White et al. 2001) for orders $p = 1$ to 4. In general, although there is an absence of pixels where the gradients of two or more variables are simultaneously high, there is practically no correlation between gradients, as shown by the low r value and the horizontal and vertical regression lines $X \rightarrow Y$ and $Y \rightarrow X$, respectively. As p increases, the maximal gradient values increase (note the increasing ranges of the axes) because of a small proportion of high-gradient points, but by and large the bulk of the cloud changes little. In particular, there is neither a positive correlation between gradients nor a negative one, or very small if at all (as in Bosking et al. 2002; Buzás et al. 2003; Hetherington and Swindale 1999; White et al. 2001; and in contrast with Das and Gilbert 1997). What is remarkable, and confirms what we said earlier, is the correlation between the gradients of OD and OR at points at or near pinwheels (which is responsible for the apparent trend in the slope of the regression lines). For $p = 1$, pinwheels (by definition characterized by a high OR gradient) lie almost exclusively in areas of zero OD gradient (i.e., away from borders). As p increases, the pinwheel population separates into 2 groups, one at near-zero OD gradient and another at high OD gradient (i.e., the OD borders), with very few pinwheels lying in between.

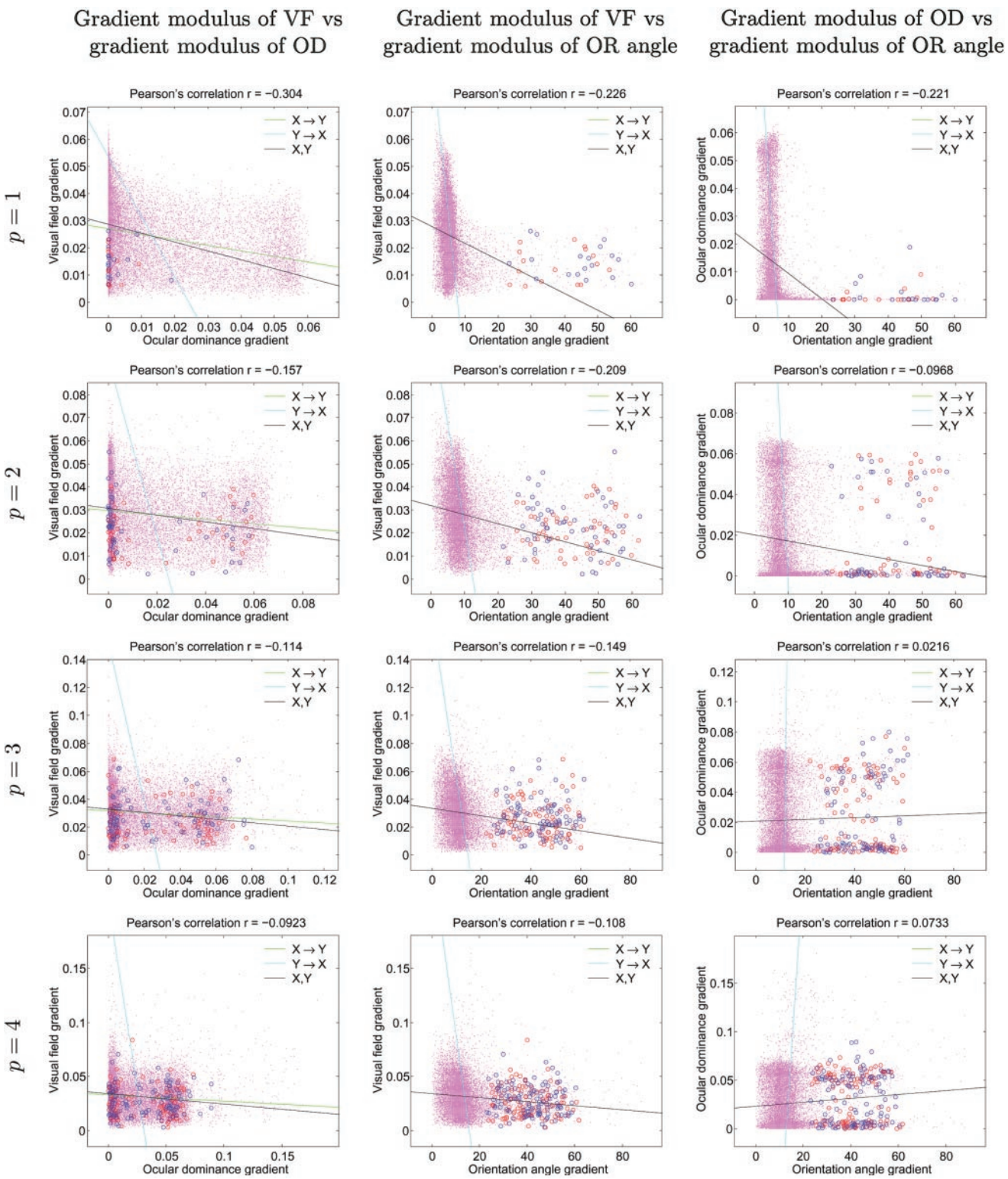
DISCUSSION

Although a dense network of short-range cortical connections is known to exist in visual cortex (e.g., Amir et al. 1993; Bosking et al. 1997; Callaway 1998; Kisvárday et al. 1997; Lund et al. 2003; Stettler et al. 2002), the precise form of the resulting intracortical interaction function is unclear. We have examined the effect of varying the form of the continuity term representing these short-range interactions in the elastic net model of visual cortical map development. In particular, we have investigated continuity terms corresponding to 1st-, 2nd-, 3rd-, and 4th-order derivative constraints on the rate of change of cortical receptive field location in a feature space consisting of visual field position, ocular dominance, and orientation preference. We have shown that this progression of derivatives corresponds to a progression of patterns of short-range intracortical interactions with increasing numbers of oscillations in the sign of the interactions. Our simulations lead us to hypothesize that the greater the number of oscillations in the continuity term, the stronger the co-location between OR pinwheels and OD borders. In particular, we predict that lateral connections oscillate only once in the regions of cat and monkey cortex that have been well characterized by optical imaging so far. These results provide a way to infer the form of the lateral connections from the observed map structure. Thus if in other species, visual cortical areas, or at other developmental stages a strong co-location between OR pinwheels and OD borders is observed, this can be interpreted in our model in terms of lateral interactions that oscillate more than once.

In fact, a recent experimental report might be consistent with this. Kisvárday et al. (2001) optically imaged the maps of OD and OR in adult cats (>4 mo) and kittens (<4 mo). They found that the tendency of pinwheels to lie near OD centers and of OD and OR maps to intersect at steep angles was hardly present in the region between areas 17 and 18, and very weak in area 18. Further, the tendency was significantly weaker for adult cats than for kittens in all 3 regions. These results suggest that the arrangement of pinwheels with respect to OD domains may change with age and cortical region. It is also worth noting that published experimental data do show a proportion, though small, of pinwheels right on OD borders (e.g., Fig. 4*A* of Hübener et al. 1997).

The growing co-location of discontinuities of OR (pinwheels) and OD (borders) and the flattening of the angle histogram with increasing p seems to be an essential feature of the elastic net. However, although the dependency of these statistics on p is generally gradual and monotonic, there seem to be 2 qualitatively different classes of lateral interactions: one for $p = 1$ and another for $p \geq 2$, with $p = 2$ often being a

transition case, as evidenced by the sharp migration of pinwheels to OD borders for $p > 1$ (Figs. 1 and 2). Another characteristic feature of the generalized elastic net is the narrowing of the stripes and the fact that the ratio of OD to OR wavelength is smaller for the 1st-order derivative than for the higher-order ones, given by Fig. 4 (*left*). In principle, this might suggest cats have cortical interaction functions of higher order than monkeys (consistent with there being less tendency toward pinwheels lying in OD centers in cats compared with monkeys). Also, this could suggest that any changes in the ratio of wavelengths in an individual between the fovea and periphery might be indicative of a change in lateral interactions over



the cortex. However, the stripe widths are also strongly determined by the correlations in the input space (Goodhill and Cimponeriu 2000) and the strength of the continuity term β, which makes it difficult to separate the individual contributions of each of these factors.

Our results show that interrelations between OD and OR maps are caused by the specific form of lateral interactions used by the elastic net. From a structural perspective, what is special about the 1st-order derivative (i.e., lateral interactions) that oscillate only once? This form was originally used in the elastic net because it is relevant to solving the Traveling Salesman Problem, that is, minimizing (squared) distances between the receptive fields in feature space of neurons that are neighboring in the cortex (Durbin and Mitchison 1990; Durbin and Willshaw 1987). This also had a flavor of wire-length minimization, although in the input feature space (Durbin and Mitchison 1990) rather than the cortical space (Koulakov and Chklovskii 2001). Durbin and Mitchison (1990) explored the effect of varying the exponent in the sum-of-distances term, but did not explore higher-order derivatives as we have done. The motivation for exploring higher-order derivatives also comes from regularization theory (Tikhonov and Arsenin 1977). Derivative-based penalties can be used to represent biological or physical constraints and so obtain good solutions to inverse problems in areas as diverse as vision (Poggio et al. 1985), geophysics (Tarantola 1987), or statistical smoothing and spline approximation (Green and Silverman 1994; Wahba 1990). The effect of the derivative order on the stripe width is theoretically understood in some of these problems (Wahba 1990; Carreira-Perpiñán and Goodhill, unpublished observations). However, at the present such a theoretical understanding is lacking for the other aspects of the 2D structure of the maps that are more subtle but occur reliably in our simulations, such as crossing angles and pinwheel locations. Note also that the different maps are independent in the continuity term, so that these effects are attributed to the combined action with the coverage term. However, the uncertainty about the definition of coverage is smaller than that of continuity (Carreira-Perpiñán and Goodhill 2002a; Swindale et al. 2000) and we may expect similar effects with slightly different definitions of coverage. From a functional perspective, the consequences of any of these different forms for the lateral interactions remain unclear.

It should be noted that different models have somewhat different ways of mapping between the form of cortical continuity they implicitly or explicitly use and a pattern of lateral interactions. For instance, in correlational models, such as that of Miller et al. (1989), the effective cortical interaction $\mathbf{I}$ maps to a pattern of lateral interactions $\mathbf{B}$ by the relation $\mathbf{I} = (\mathbf{1} - \mathbf{B})^{-1}$ (where $\mathbf{I}$ and $\mathbf{B}$ are matrices and $\mathbf{1}$ represents the identity matrix; Miller and Stryker 1990; Miller et al. 1989). In this model, $\mathbf{I}$ values of Mexican-hat form generally thus correspond to patterns of lateral interactions with many oscillations, and when the joint development of OR and OD is simulated with this model, neither the orthogonality between the 2 columnar systems nor the positioning of OR pinwheels at the center of OD columns is robustly reproduced (Erwin and Miller 1998; see also Piepenbrock et al. 1997). Mathematical relationships between correlational and elastic-net–type models are discussed by Dayan (1993) and Yuille et al. (1996).

In general, we find a substantially weaker relationship in our simulations between the rate of change of different input features with cortical position than has generally been thought to occur in the elastic net (Durbin and Mitchison 1990). Also, by simulating tangential penetrations across pinwheels (data not shown) we do not generally find a jump in the VF position (as confirmed by the fact that pinwheels lie in areas of low to middle VF gradient in Fig. 5). The diffuse clouds of points in Fig. 5 demonstrate that many types of relationships between gradients of variables coexist in the simulated cortex. The elastic net (or dimension-reduction models in general) is not best characterized as producing purely negatively correlated representations where, in general, when one feature varies substantially, the remaining ones vary only slightly (Durbin and Mitchison 1990). Das and Gilbert (1997) claimed that the rate of change of orientation preference is positively correlated with that of visual field position in the cat, although this has not been reproduced by other investigators (cat: Buzás et al. 2003; Hetherington and Swindale 1999; ferret: White et al. 2001; tree shrew: Bosking et al. 2002). It is interesting to note that a model specifically formulated to obtain this result (Mitchison and Swindale 1999) yielded as a by-product pinwheels on OD borders (corresponding to a positive correlation between the rate of change of OR and OD), which is contrary to what is observed in optical imaging images. Our results show that, as the derivative order increases, the discontinuities of OD and OR increasingly tend to co-locate, but the gradients of OR and VF do not correlate.

However, it is important to note that the methodology of experimental studies is practically constrained to be quite different from the complete analysis of rates of changes between feature variables that is possible in theoretical models (Fig. 5). Our gradients are, up to the small discretization error induced by the grid, "true" gradients. Gradients computed from a set of electrode penetrations cannot be considered true gradients for 3 reasons: first, they are really directional derivatives along the line joining 2 sites (the gradient equation takes into account all directions); second, the separation between sites is usually so large (>100 μm, or about 10% of the OR wavelength) that the discretization error becomes very large

FIG. 5. Pairwise scatterplots of the gradient moduli of visual field, ocular dominance, and orientation angle for derivative orders $p = 1$ to 4, for the maps of Fig. 1. Total number of points in each cloud is $128 \times 128 = 16{,}384$, equal to the number M of points in the net. Pinwheels are marked by red (positive) and blue (negative) circles. Units of the axes are # per pixel, where # is degrees for the orientation angle and arbitrary for the visual field and OD (i.e., as given by the training set). Note the different scales in the axes. Lines shown are the regression of Y on X ($X \rightarrow Y$), of X on Y ($Y \rightarrow X$), and the principal component (X, Y). Clouds of points are diffuse, indicating that all combinations of gradients are abundantly represented, also evidenced by the low correlation coefficients. In general, the clouds appear compressed because of a rescaling of the axes needed to accommodate a few points of very high gradient (such as pinwheels). Apparent trend in the slope of the regression lines as p increases is caused by the increase in the number of pinwheels and their separation into 2 groups: right at OD borders (high OD gradient) and away from them (near-zero OD gradient). These 2 groups correspond to the two peaks at 0 and approximately λ(OD)/4, respectively, in the histogram of pinwheel-to-closest-OD-border distance of Fig. 2 (*center column*).

and can corrupt any local correlation or lack of it; third, the number of sites is very small (a few dozens compared with $M = 16{,}384$ in our simulations). The significance of the location of pinwheels in areas of low to medium VF gradients with respect to coverage uniformity (Bosking et al. 2002; Das and Gilbert 1997) remains unclear.

ACKNOWLEDGMENTS

We are very grateful to P. Dayan for helpful discussions.

Present address of M. Á. Carreira-Perpiñán: Dept. of Computer Science, University of Toronto, 6 King's Rd., Toronto, ON M5S 3H5, Canada (E-mail: miguel@cs.toronto.edu).

REFERENCES

Amir Y, Harel M, and Malach R. Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J Comp Neurol* 334: 19–46, 1993.

Ball P. *The Self-Made Tapestry. Pattern Formation in Nature*. Oxford, UK: Oxford Univ. Press, 1999.

Bartfeld E and Grinvald A. Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proc Natl Acad Sci USA* 89: 11905–11909, 1992.

Bosking WH, Crowley JC, and Fitzpatrick D. Spatial coding of position and orientation in primary visual cortex. *Nat Neurosci* 5: 874–882, 2002.

Bosking WH, Zhang Y, Schofield B, and Fitzpatrick D. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J Neurosci* 17: 2112–2127, 1997.

Buzás P, Volgushev M, Eysel UT, and Kisvárday ZF. Independence of visuotopic representation and orientation map in the visual cortex of the cat. *Eur J Neurosci* 18: 957–968, 2003.

Callaway EM. Local circuits in primary visual cortex of the macaque monkey. *Annu Rev Neurosci* 21: 47–74, 1998.

Carreira-Perpiñán MÁ and Goodhill GJ. Are visual cortex maps optimized for coverage? *Neural Comput* 14: 1545–1560, 2002a.

Carreira-Perpiñán MÁ and Goodhill GJ. Development of columnar structures in visual cortex. In: *Computational Neuroanatomy*: *Principles and Methods*, edited by Ascoli GA. Totowa, NJ: Humana Press, 2002b, chap. 15, p. 337–357.

Cover TM and Thomas JA. *Elements of Information Theory*. New York: Wiley, 1991.

Crair MC, Ruthazer ES, Gillespie DC, and Stryker MP. Ocular dominance peaks at pinwheel center singularities of the orientation map in cat visual cortex. *J Neurophysiol* 77: 3381–3385, 1997.

Das A and Gilbert CD. Distortions of visuotopic map match orientation singularities in primary visual cortex. *Nature* 387: 594–598, 1997.

Dayan P. Arbitrary elastic topologies and ocular dominance. *Neural Comput* 5: 392–401, 1993.

Dragoi V and Sur M. Dynamic properties of recurrent inhibition in primary visual cortex: contrast and orientation dependence of contextual effects. *J Neurophysiol* 83: 1019–1030, 2000.

Durbin R and Mitchison G. A dimension reduction framework for understanding cortical maps. *Nature* 343: 644–647, 1990.

Durbin R, Szeliski R, and Yuille A. An analysis of the elastic net approach to the traveling salesman problem. *Neural Comput* 1: 348–358, 1989.

Durbin R and Willshaw D. An analogue approach to the traveling salesman problem using an elastic net method. *Nature* 326: 689–691, 1987.

Ernst UA, Pawelzik KR, Sahar-Pikielny C, and Tsodyks MV. Intracortical origin of visual maps. *Nat Neurosci* 4: 431–436, 2001.

Erwin E and Miller KD. Correlation-based development of ocularly matched orientation and ocular dominance maps: determination of required input activities. *J Neurosci* 18: 9870–9895, 1998.

Erwin E, Obermayer K, and Schulten K. Models of orientation and ocular dominance columns in the visual cortex: a critical comparison. *Neural Comput* 7: 425–468, 1995.

Fagiolini M, Fritschy JM, Low K, Mohler H, Rudolph U, and Hensch TK. Specific GABAA circuits for visual cortical plasticity. *Science* 303: 1681–1683, 2004.

Goodhill GJ. Topography and ocular dominance: a model exploring positive correlations. *Biol Cybern* 69: 109–118, 1993.

Goodhill GJ and Cimponeriu A. Analysis of the elastic net applied to the formation of ocular dominance and orientation columns. *Netw Comput Neural Syst* 11: 153–168, 2000.

Goodhill GJ and Willshaw DJ. Application of the elastic net algorithm to the formation of ocular dominance stripes. *Netw Comput Neural Syst* 1: 41–59, 1990.

Green PJ and Silverman BW. *Nonparametric Regression and Generalized Linear Models*: *A Roughness Penalty Approach*. New York: Chapman & Hall, 1994.

Hata Y, Tsumoto T, Sato H, and Tamura H. Horizontal interactions between visual cortical neurones studied by cross-correlation analysis in the cat. *J Physiol* 441: 593–614, 1991.

Hensch TK and Stryker MP. Columnar architecture sculpted by GABA circuits in developing cat visual cortex. *Science* 303: 1678–1681, 2004.

Hetherington PA and Swindale NV. Receptive field and orientation scatter studied by tetrode recordings in cat area 17. *Vis Neurosci* 16: 637–652, 1999.

Hübener M, Shoham D, Grinvald A, and Bonhoeffer T. Spatial relationships among three columnar systems in cat area 17. *J Neurosci* 17: 9270–9284, 1997.

Kang K, Shelley M, and Sompolinsky H. Mexican hats and pinwheels in visual cortex. *Proc Natl Acad Sci USA* 100: 2848–2853, 2003.

Kim DS, Matsuda Y, Ohki K, Ajima A, and Tanaka S. Geometrical and topological relationships between multiple functional maps in cat primary visual cortex. *Neuroreport* 10: 2515–2522, 1999.

Kisvárday ZF, Tóth É, Rausch M, and Eysel UT. Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cereb Cortex* 7: 605–618, 1997.

Kisvárday ZF, Womelsdorf T, Buzás P, and Eysel UT. Variability of pinwheel locations in ocular dominance maps according to area and age. *Soc Neurosci Abstr* 27: 285.4, 2001.

Koulakov AA and Chklovskii DB. Orientation preference patterns in mammalian visual cortex: a wire length minimization approach. *Neuron* 29: 519–527, 2001.

Löwel S, Bischof HJ, Leutenecker B, and Singer W. Topographic relations between ocular dominance and orientation columns in the cat striate cortex. *Exp Brain Res* 71: 33–46, 1988.

Lund JS, Angelucci A, and Bressloff PC. Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cereb Cortex* 13: 15–24, 2003.

Meinhardt H. *Models of Biological Pattern Formation*. New York: Academic Press, 1982.

Miller KD, Keller JB, and Stryker MP. Ocular dominance column development: analysis and simulation. *Science* 245: 605–615, 1989.

Miller KD and Stryker MP. The development of ocular dominance columns: mechanisms and models. In: *Connectionist Modeling and Brain Function: The Developing Interface*, edited by Hanson SJ and Olson CR. Cambridge, MA: MIT Press, 1990, chap. 9, p. 255–350.

Mitchison GJ and Swindale NV. Can Hebbian volume learning explain discontinuities in cortical maps? *Neural Comput* 11: 1519–1526, 1999.

Müller T, Stetter M, Hübener M, Sengpiel F, Bonhoeffer T, Godecke I, Chapman B, Löwel S, and Obermayer K. An analysis of orientation and ocular dominance patterns in the visual cortex of cats and ferrets. *Neural Comput* 12: 2573–2595, 2000.

Obermayer K and Blasdel GG. Geometry of orientation and ocular dominance columns in monkey striate cortex. *J Neurosci* 13: 4114–4129, 1993.

Obermayer K and Blasdel GG. Singularities in primate orientation maps. *Neural Comput* 9: 555–575, 1997.

Piepenbrock C, Ritter H, and Obermayer K. The joint development of orientation and ocular dominance: role of constraints. *Neural Comput* 9: 959–970, 1997.

Poggio T, Torre V, and Koch C. Computational vision and regularization theory. *Nature* 317: 314–319, 1985.

Roerig B and Chen B. Relationships of local inhibitory and excitatory circuits to orientation preference maps in ferret visual cortex. *Cereb Cortex* 12: 187–198, 2002.

Rose K. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proc IEEE* 86: 2210–2239, 1998.

Schuett S, Bonhoeffer T, and Hübener M. Mapping retinotopic structure in mouse visual cortex with optical imaging. *J Neurosci* 22: 6549–6559, 2002.

Sirosh J and Miikkulainen R. Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Comput* 9: 577–594, 1997.

Somers DC, Todorov EV, Siapas AG, Toth LJ, Kim DS, and Sur M. A local circuit approach to understanding integration of long-range inputs in primary visual cortex. *Cereb Cortex* 8: 204–217, 1998.

Stettler DD, Das A, Bennett J, and Gilbert CD. Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron* 36: 739–750, 2002.

Sur M and Leamey CA. Development and plasticity of cortical areas and networks. *Nat Rev Neurosci* 2: 251–262, 2001.

Swindale NV. Coverage and the design of striate cortex. *Biol Cybern* 65: 415–424, 1991.

Swindale NV. The development of topography in the visual cortex: a review of models. *Netw Comput Neural Syst* 7: 161–247, 1996.

Swindale NV, Shoham D, Grinvald A, Bonhoeffer T, and Hübener M. Visual cortex maps are optimised for uniform coverage. *Nat Neurosci* 3: 822–826, 2000.

Tarantola A. *Inverse Problem Theory*: *Methods for Data Fitting and Model Parameter Estimation*. Amsterdam: Elsevier Science, 1987.

Tikhonov AN and Arsenin VY. *Solutions of Ill-Posed Problems*. New York: Wiley, 1977.

Turing AM. The chemical basis of morphogenesis. *Proc R Soc Lond B Biol Sci* 237: 37–72, 1952.

von der Malsburg C. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* 14: 85–100, 1973.

Wahba G. *Spline Models for Observational Data*. Philadelphia, PA: SIAM, 1990.

White LE, Bosking WH, and Fitzpatrick D. Consistent mapping of orientation preference across irregular functional domains in ferret visual cortex. *Vis Neurosci* 18: 65–76, 2001.

Willshaw DJ and von der Malsburg C. How patterned neural connections can be set up by self-organization. *Proc R Soc Lond B Biol Sci* 194: 431–445, 1976.

Yuille AL, Kolodny JA, and Lee CW. Dimension reduction, generalized deformable models and the development of ocularity and orientation. *Neural Networks* 9: 309–319, 1996.