NOTE Communicated by Klaus Obermayer

Are Visual Cortex Maps Optimized for Coverage?

Miguel Á. Carreira-Perpiñán
miguel@cns.georgetown.edu
Geoffrey J. Goodhill
geoff@georgetown.edu
Department of Neuroscience,
Georgetown University Medical Center, Washington, D.C. 20007, U.S.A.

The elegant regularity of maps of variables such as ocular dominance, orientation, and spatial frequency in primary visual cortex has prompted many people to suggest that their structure could be explained by an optimization principle. Up to now, the standard way to test this hypothesis has been to generate artificial maps by optimizing a hypothesized objective function and then to compare these artificial maps with real maps using a variety of quantitative criteria. If the artificial maps are similar to the real maps, this provides some evidence that the real cortex may be optimizing a similar function to the one hypothesized. Recently, a more direct method has been proposed for testing whether real maps represent local optima of an objective function (Swindale, Shoham, Grinvald, Bonhoeffer, & Hübener, 2000). In this approach, the value of the hypothesized function is calculated for a real map, and then the real map is perturbed in certain ways and the function recalculated. If each of these perturbations leads to a worsening of the function, it is tempting to conclude that the real map is quite likely to represent a local optimum of that function. In this article, we argue that such perturbation results provide only weak evidence in favor of the optimization hypothesis.

1 Introduction

Neurons in visual cortex respond to several kinds of visual stimuli, the best studied of which include position in visual field, eye of origin, and orientation, direction, and spatial frequency of a grating. The pattern of preferred stimulus values over the whole visual cortex for each kind of stimulus is called a (visual) cortical map. Thus, maps of visual field position, ocular dominance, orientation, and so forth coexist on the same neural substrate. Given that these maps show a highly organised spatial structure, the question arises of what underlying principles explain these maps. Two such principles are *coverage uniformity*, or completeness, and *continuity*, or similarity (Hubel & Wiesel, 1977). Coverage uniformity means that each combination

of stimuli values (e.g., any orientation in any visual field location of either eye) has equal representation in the cortex; completeness means that any combination of stimuli values is represented somewhere in cortex. Thus, coverage uniformity implies completeness (disregarding the trivial case of a cortex uniformly nonresponsive to stimuli), but not vice versa, since it is possible to have over- and underrepresented stimuli values (in addition, it is not practically possible to represent all values of a continuous higher-dimensional stimulus space with a continuous two-dimensional cortex). A useful middle ground is to consider that the set of stimulus values represented by the cortex be roughly uniformly scattered in stimulus space. A common qualitative definition of continuity is that neurons that are physically close in cortex tend to have similar stimulus preferences; this can be motivated in terms of economy of cortical wiring (Durbin & Mitchison, 1990).

Coverage and continuity compete with each other. If, say, retinotopy and preferred orientation vary slowly from neuron to neuron, sizable visual field regions will lack some orientations. If neurons' preferred stimuli values are scattered like a salt and pepper mixture, continuity is lost. The striped structure of several of the maps can be seen as a compromise between these two extremes. An early model based on these principles is the ice cube model of Hubel and Wiesel (1977), where stripes of ocular dominance run orthogonally to stripes of orientation and all combinations of eye and orientation preference are represented within a cortical region smaller than a cortical point image (the collection of neurons whose receptive fields contain a given visual field location). The competition can be explained in a dimension-reduction framework, where a two-dimensional cortical sheet twists in a higher-dimensional stimulus space to cover it as uniformly as possible while minimizing some measure of continuity. Optimization models based on such principles produce maps with a quantitatively good match to the observed phenomenology of cortical maps, including the striped structure of ocular dominance and orientation columns with appropriate periodicity and interrelations (Erwin, Obermayer, & Schulten, 1995; Swindale, 1996).

However, a more direct approach to test the validity of such optimization models would be to calculate the value of the objective function for a real cortical map and then determine by perturbation whether this represents a local optimum. Such an approach has recently been proposed by Swindale, Shoham, Grinvald, Bonhoeffer, & Hübener (2000). Although the results they presented are consistent with the hypothesis that real maps are optimized for a particular function measuring coverage, here we argue that these results offer only weak evidence in favor of the hypothesis.

2 The Coverage Measure

Consider a resolution-dependent representation of a cortical map defined as a two-dimensional array of vector values of the stimulus variables of interest. Each position (i, j) in the array represents an ideal cortical cell;

call $\mathcal{C}$ the set of all such cortical positions. There is a vector of stimulus values $\boldsymbol{\mu}_{ij}$ associated with each cortical position (i, j); stimulus variables considered by Swindale et al. (2000) are the retinotopic position (or receptive field center in the visual field) (x, y) in degrees, the preferred orientation $\theta \in [0^\circ, 180^\circ)$, the ocular dominance n (-1: left eye, $+1$: right eye), and the spatial frequency $m \in \{-1, 1\}$. Therefore, $\boldsymbol{\mu}_{ij} \stackrel{\text{def}}{=} (n_{ij}, m_{ij}, \theta_{ij}, x_{ij}, y_{ij})$ for $(i, j) \in \mathcal{C}$ can be considered a generalized receptive field center; a receptive field would then be defined by a function sitting on the receptive field center and monotonically decreasing away from it (see below). The collection $\mathbf{M} \stackrel{\text{def}}{=} \{\boldsymbol{\mu}_{ij}\}_{(i,j)\in\mathcal{C}}$ of such receptive field centers, together with the two-dimensional ordering of cortical positions in $\mathcal{C}$, defines the cortical map.

A mathematically convenient way of representing the trade-off between the goals of attaining uniform coverage and respecting the constraints of cortical wiring is to assume that cortical maps maximize a function

$$\mathcal{F}(\mathbf{M}) \stackrel{\text{def}}{=} \mathcal{C}(\mathbf{M}) + \lambda\mathcal{R}(\mathbf{M}), \tag{2.1}$$

where $\mathcal{C}$ is a measure of the uniformity of coverage, $\mathcal{R}$ is a measure of the continuity, and $\lambda > 0$ specifies the relative weight of $\mathcal{R}$ with respect to $\mathcal{C}$. We assume that maximizing either $\mathcal{C}$ or $\mathcal{R}$ separately does not lead to a maximum of $\mathcal{F}$ and therefore that maxima of $\mathcal{F}$ imply compromise values of $\mathcal{C}$ and $\mathcal{R}$. The exact form of the combination of $\mathcal{C}$ and $\mathcal{R}$ of equation 2.1 (a weighted sum) need not be biologically correct, but for the purposes of embodying the competition between $\mathcal{C}$ and $\mathcal{R}$, it is sufficient.

Swindale (1991) introduced the following mathematical definition of coverage. Given an arbitrary stimulus $\mathbf{v}$, the total amount of cortical activity that it produces is defined as

$$A(\mathbf{v}) \stackrel{\text{def}}{=} \sum_{(i,j)\in\mathcal{C}} f(\mathbf{v} - \boldsymbol{\mu}_{ij}), \tag{2.2}$$

where f is the (generalized) receptive field of cortical location (i, j), assumed translationally invariant (so it depends on only the difference of stimulus $\mathbf{v}$ and generalized receptive field center $\boldsymbol{\mu}_{ij}$); f is taken as a product of functions: gaussian for orientation and retinotopic position (with widths derived from biological estimates of tuning curves)[1] and delta for ocular dominance and spatial frequency. A is calculated for a regular grid in stimulus space, which is assumed to be a representative set of stimulus values. The measure

[1] Strictly, the receptive field size depends on the location of the stimulus in the visual field and adapts to the surround (e.g., with contrast). However, extending the coverage definition to account for this is difficult. Therefore, in common with Swindale et al. (2000), we will consider fixed receptive field sizes.

of coverage uniformity is finally obtained as

$$c' \stackrel{\text{def}}{=} \frac{\text{stdev}\{A\}}{\text{mean}\{A\}}, \tag{2.3}$$

that is, the magnitude of the normalized dispersion of the total activity A in the stimulus space. Intuitively, c' will be large when A takes different values for different stimuli and zero if A has the same value independent of the stimulus. Thus, it is a measure of lack of coverage uniformity, and we could define $\mathcal{C} \stackrel{\text{def}}{=} -c'$. Equation 2.2 can be seen as a generalization of the fitness term of the elastic net objective function (Durbin, Szeliski, & Yuille, 1989). $\mathcal{R}$ is the combined effect of several factors, none of which is fully understood, and so it is hard to write down a functional form for it confidently.

3 Determining Map Optimality via Perturbations

If suitable functions $\mathcal{C}$ and $\mathcal{R}$ are defined, the mathematical procedure to determine whether a given map $\mathbf{M} = \{\boldsymbol{\mu}_{ij}\}_{(i,j)\in\mathcal{C}}$ is a (local) maximum of $\mathcal{F}$ is to check that the gradient of $\mathcal{F}$ at $\mathbf{M}$ is zero and the Hessian of $\mathcal{F}$ at $\mathbf{M}$ is negative definite (or negative semidefinite). However, there are two problems with this. First, $\mathcal{C}$ is obtained in an approximate way[2] using a sample of the stimulus distribution, and so the numerical accuracy of the gradient and Hessian will be affected by a discretization error, particularly if the sample is coarse and symmetric. But second and crucially, even if we commit ourselves to a given (approximated) mathematical definition of $\mathcal{C}$ such as $-c'$, we still do not have a suitable definition of $\mathcal{R}$.

Given the difficulties in the definition of $\mathcal{R}$ and the mathematical treatment of $\mathcal{C}$, the goal of Swindale et al. (2000) was less ambitious: to check whether the maps are at local optima of $\mathcal{C}$ by examining the effect on $\mathcal{C}$ of a fairly small set of perturbations of the maps that hopefully would not affect $\mathcal{R}$, however the latter is defined. That is, they argued that although we do not know what $\mathcal{R}$ is exactly, we may be able to determine what perturbations of a map should leave $\mathcal{R}$ unaffected. Specifically, they suggested rigid motion perturbations (horizontal translations, 180 degree rotations, and horizon-

[2] Although we have a mathematical definition of $\mathcal{C} = -c'$ in terms of $\mathbf{M}$ (via the intermediate definition of $A(\mathbf{v}; \mathbf{M})$), it is not possible to obtain $\mathcal{C}$ as an explicit function of $\mathbf{M}$ for a given distribution of the stimulus $\mathbf{v}$ (e.g., uniform) since one cannot analytically determine the distribution of A. Thus, $\mathcal{C}(\mathbf{M})$ must be approximated with a sample of the stimulus distribution. Swindale et al. (2000) computed the total cortical activity A for all combinations of $n \in \{-1, 1\}$, $m \in \{-1, 1\}$, $\theta \in \{0°, 30°, 60°, 90°, 120°, 150°\}$, $x \in \{1, 5, 9, \ldots, i_{\max}\}$, and $y \in \{1, 5, 9, \ldots, j_{\max}\}$, where $i_{\max}$ and $j_{\max}$ are the size in pixels of the rectangular map, amounting to about 14,000 stimuli. This is a coarse and symmetric sample of θ, x, and y. Drawing a finer random sample from a uniform distribution in stimulus space could avoid potential artifactual estimates and could also be used to control whether different samples lead to essentially the same value of A.

tal and vertical flips) applied separately to each individual map (discussed further in section 3.3). If such perturbations unambiguously worsened coverage uniformity for biologically observed maps of developed animals, it would be tempting to conclude that such maps are local maxima of both $\mathcal{C}$ and $\mathcal{F}$. To test this idea, Swindale et al. (2000) used empirical maps of ocular dominance, orientation, and spatial frequency obtained simultaneously in area 17 of the cat using standard optical imaging methods for young animals. After some preprocessing (including smoothing, necessary to remove noise), rectangular regions of about 5×2.5 mm (approximately 140×70 pixels) were obtained in which each pixel has associated values of ocular dominance in $\{-1, 1\}$, orientation in degrees in $[0, 180)$, and spatial frequency in $\{-1, 1\}$. Since optical imaging provides no information about topography, they chose to make the retinotopic map linear, that is, perfectly topographic,[3] which makes coverage uniform by definition along the retinotopic variables x, y. Swindale et al. (2000) then computed the variation of c' for a range of rigid-motion perturbations and found that coverage became less uniform for most of the perturbations described, often as an increasing function of the size of the perturbation (notably for horizontal shifts).

The lack of negative results in the perturbation simulations led Swindale et al. (2000) to argue that such maps are local maxima of both $\mathcal{C}$ and $\mathcal{F}$—or as Das (2000) put it in an associated article, "Real experimentally obtained maps from V1 are indeed optimally arranged with respect to each other such that any departure from the real maps worsens the coverage." However, there are several reasons to question whether this really follows from the results presented by Swindale et al.

3.1 Incompleteness of the Perturbation Set. The set of perturbations used by Swindale et al. (2000) does not include all possible perturbations that would leave a certain continuity function $\mathcal{R}$ unchanged. Assume that all stimulus variables are continuous, and call D the number of (independent) such variables, that is, the number of scalar variables in $\mathbf{M} = \{\boldsymbol{\mu}_{ij}\}_{(i,j)\in\mathcal{C}}$. This is a large number: for a rectangular map of 140×70 with 5 stimulus variables, D is $49,000$, and it could even be infinite if one considers a nonparametric representation of the map (in which case we would have a variational problem). Then an elementary perturbation of a map $\mathbf{M}$ that does not alter $\mathcal{R}$ will result in a perturbed map lying at an ϵ-distance from $\mathbf{M}$ on the manifold $\mathcal{R}(\mathbf{M}) \stackrel{\text{def}}{=} \{\mathbf{N}\colon\ \mathcal{R}(\mathbf{N}) = \mathcal{R}(\mathbf{M})\}$ (see Figure 1). Such a manifold will have dimension $D-1$ (or less, if some variables are dependent). Therefore, if $D > 2$, the number of different directions in which to perturb the map is infinite. In other words, proving that $\mathbf{M}$ is a maximum of $\mathcal{C}$ for fixed $\mathcal{R}$ requires proving

[3] The effect on the coverage estimates of assuming that the retinotopic map is strictly topographic is likely to be considerable. Besides, this assumption may not be true in reality (e.g., see Das & Gilbert, 1997).

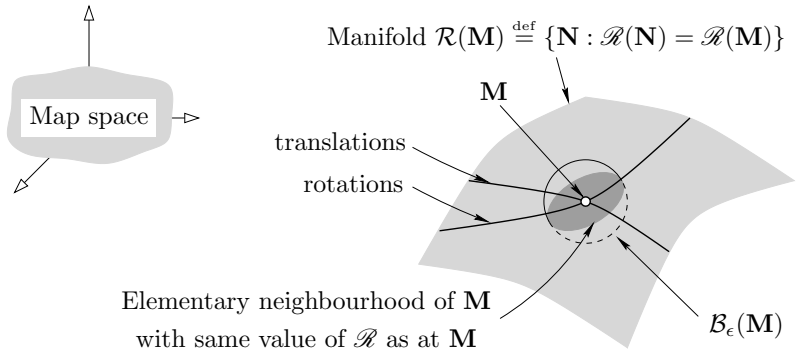


Figure 1: Illustration of an elementary neighborhood of $\mathbf{M}$ contained in the manifold $\mathcal{R}(\mathbf{M}) \stackrel{\text{def}}{=} \{\mathbf{N}\text{: } \mathcal{R}(\mathbf{N}) = \mathcal{R}(\mathbf{M})\}$. In this example, the map space is three-dimensional, and the manifold is two-dimensional. The thick lines indicate manifolds along which some specific classes of perturbations leave $\mathcal{R}$ constant. These are just a subset of all perturbations that leave $\mathcal{R}$ constant (the dark-shaded neighborhood). $\mathcal{B}_\epsilon(\mathbf{M})$ is a ball of radius ϵ centered at $\mathbf{M}$.

that every perturbed map in an elementary neighborhood of $\mathbf{M}$ contained in the mentioned manifold has a lower value of $\mathcal{C}$. Such a neighborhood can be specifically defined as the intersection of a ball $\mathcal{B}_\epsilon(\mathbf{M})$ of small radius $\epsilon > 0$ and the manifold $\mathcal{R}(\mathbf{M})$, as shown in Figure 1, and contains an infinite number of maps. Therefore, a procedure based on trying a finite number of different perturbations can never prove the statement, although it can disprove it by finding one such perturbation that increases $\mathcal{C}$ (assuming it is possible to implement an elementary perturbation numerically). No matter how many perturbations we try that decrease $\mathcal{C}$, we can never be sure that the rest of them will as well. Even if a candidate map passes a seemingly convincingly large number of perturbations, there are still many more perturbations to test, and there are many other candidates that would pass those perturbations too. Therefore, a "mathematically complete exploration of a full range of distortions" (Das, 2000) is an unattainable goal.

The elementary perturbations we have mentioned include not just (vanishingly small) rigid motions in all possible directions, but also different classes of perturbations, such as "rubber-sheet" distortions of the map. Both Swindale et al. (2000) and Das (2000) admit the possibility that rubber-sheet distortions exist that improve coverage uniformity while leaving $\mathcal{R}$ unchanged. However, Swindale et al. argue that since $\mathcal{R}$ is not properly defined, a given distortion that leaves unchanged a function $\mathcal{R}_1$ defined in some way would likely change a function $\mathcal{R}_2$ defined in a different way. Besides the fact that the same argument could be equally applied to the definition of coverage uniformity (a distortion decreasing coverage uniformity

as c' does might increase coverage uniformity under a different definition), our argument remains, since for any given definition of $\mathcal{R}$, there potentially exist distortions that could increase $\mathcal{C}$, that is, the neighborhood defined earlier depends on the chosen definition of $\mathcal{R}$, but it always exists. These issues are illustrated more concretely in Figure 2. This presents an example of a very simplified version of the mapping problem investigated by Swindale et al. (2000) akin to a one-dimensional ocular dominance problem (Goodhill & Willshaw, 1990). A particular map is hypothesized to be optimal and is then perturbed. We show that rigid shifts of this map decrease $\mathcal{C} = -c'$ as a monotonically decreasing function of the size of the shift. However, we also show that rubber-sheet perturbations of this map that leave $\mathcal{R}$ unchanged can *increase* $\mathcal{C}$.

3.1.1 Checking for Stationary Points. Could one at least determine whether the maps are at a stationary point of $\mathcal{C}$ for fixed $\mathcal{R}$, that is, whether the gradient of $\mathcal{C}$ at the map $\mathbf{M}$ is zero in the direction tangent to the manifold $\mathcal{R}(\mathbf{M})$? A numerical approximation to the gradient of a function f of D variables can be computed by finite differences by using small perturbations along D linearly independent directions[4]—for example, along the coordinate axes (with unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_D$), by computing

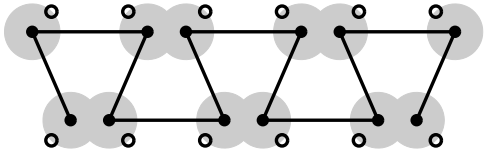
$$\frac{\partial f}{\partial x_d} \approx \frac{f(\mathbf{x} + \epsilon \mathbf{e}_d) - f(\mathbf{x})}{\epsilon}.$$

This would require f to be computed at $\mathbf{x} + \epsilon \mathbf{e}_d$ for $d = 1, \ldots, D$, that is, D component-wise small perturbations (for comparison, the shifts of Swindale et al., 2000, would amount to only six perturbation dimensions in a space of $D = 49,000$). Whether $\nabla f(\mathbf{x}) = \mathbf{0}$ could then be determined, at least up to some numerical threshold (which may not be a straightforward matter), and thus whether $\mathbf{x}$ is a stationary point of f. However, $\nabla f(\mathbf{x}) = \mathbf{0}$ is true for saddle points as well as optima, and high-dimensional multivariate functions that have many optima typically have far more saddle points (see the appendix).[5] Figure 3 shows an example of a function of two variables with a stationary point at the origin that is neither a maximum nor a minimum, but perturbing a point at the origin along many directions will result in a lower value of the function. Thus, knowing that the gradient is zero and finding that the high-dimensional function decreases along a few directions in no way guarantees that it is at a maximum.

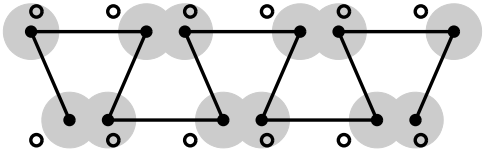
[4] For this and other statements in this section, see a text on optimization, e.g., Nocedal and Wright (1999).

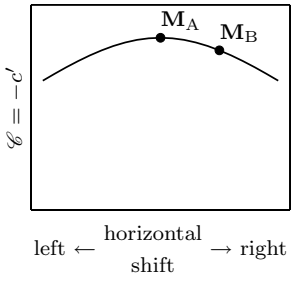
[5] That the coverage and continuity functions must have many local optima follows from symmetry considerations and from the fact that, assuming the maps are indeed optimal, while the maps of any two normal animals (of the same species) are qualitatively similar to each other, no two animals have the same map.

(A)

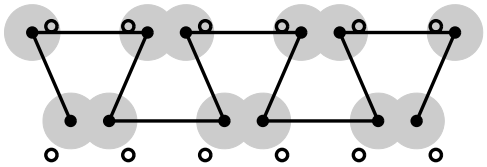


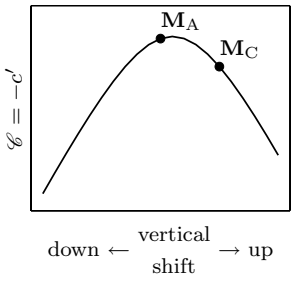
(B)



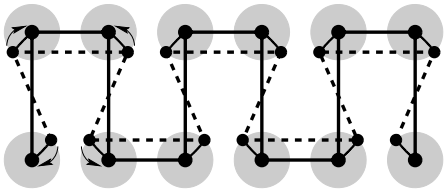

$\mathbf{M}_\mathrm{A}$
$\mathbf{M}_\mathrm{B}$
$\mathscr{C} = -c'$
left ← horizontal shift → right

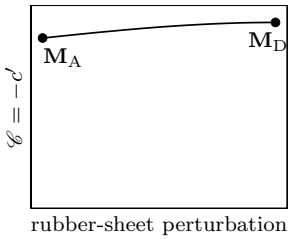
(C)




$\mathbf{M}_\mathrm{A}$
$\mathbf{M}_\mathrm{C}$
$\mathscr{C} = -c'$
down ← vertical shift → up

(D)




$\mathbf{M}_\mathrm{A}$
$\mathbf{M}_\mathrm{D}$
$\mathscr{C} = -c'$
rubber-sheet perturbation

To determine whether the stationary point is a maximum (say), one would need to compute a numerical approximation of the Hessian (the matrix of second-order derivatives) and check that it is negative definite (or positive definite, for a minimum)—that all its eigenvalues are strictly negative. This is now a much harder numerical problem than determining whether $\nabla f(\mathbf{x}) = \mathbf{0}$: estimating the Hessian requires $\mathcal{O}(D^2)$ perturbations, rather than $\mathcal{O}(D)$ as for the gradient, and even if it could be estimated, the real problem is then determining whether its eigenvalues are negative (computationally an $\mathcal{O}(D^3)$ problem). This is very difficult because it is well known that the Hessian of a function of many variables is likely to be ill conditioned: the ratio of the smallest to largest eigenvalue (in absolute value)

Figure 2: *Facing page*. A nonoptimal map that shows a systematic worsening of coverage uniformity on shifts but a systematic improvement on rubber-sheet perturbations. In this thought experiment, reminiscent of an elastic net (Durbin & Willshaw, 1987), the array of empty circles represents a uniform sample in a two-dimensional stimulus space (e.g., the horizontal axis could be a retinotopic variable and the vertical axis the ocular dominance as in Goodhill & Willshaw, 1990). The string of filled circles (receptive field centers) represents a one-dimensional map that tries to cover the stimuli as much and as uniformly as possible (measured by $\mathcal{C} \stackrel{\text{def}}{=} -c'$ as in section 2) while respecting the map continuity as much as possible (here defined as the sum $\mathcal{R}$ of the lengths of the individual segments). To compute c', in equation 2.2, a gaussian kernel f with a standard deviation equal to twice the radius of the shaded disks was used. Map B is the result of rigidly shifting map A to the right: it has the same length $\mathcal{R}$ as map A but a lower value of $\mathcal{C}$. In general, rigidly shifting map A horizontally systematically decreases $\mathcal{C}$ and results in an inverted-U curve for $\mathcal{C}$, wrongly suggesting that map A is an optimum. The same happens for vertical shifts, as in map C (although, for this particular example, map A is slightly off the maximum of $\mathcal{C}$). However, map A can be stretched and compressed symmetrically (a rubber-sheet transformation) to reach map D, keeping its length $\mathcal{R}$ constant. Thus, map A is not optimal, and there is a continuous path inside the manifold of constant $\mathcal{R}$ that monotonically increases $\mathcal{C}$ until map D is reached, as shown in map D. Whether map A is a saddle point of $\mathcal{F}(\mathbf{M}) = \mathcal{C}(\mathbf{M}) + \lambda\mathcal{R}(\mathbf{M})$ or lies on an inclined ridge depends on the actual values of λ and the parameters of $\mathcal{C}$ and $\mathcal{R}$. In all graphs in the right column, the vertical scale is the same; the steepness of the curves may be increased or decreased by changing the gaussian kernel width, and the curves were computed with a Matlab program from the actual data and definitions of $\mathcal{C}$ and $\mathcal{R}$. Note that if one considers periodic boundary conditions in the horizontal axis, by symmetry c' becomes exactly 0 for both maps A and D, even though intuitively map D is better than map A. This is a shortcoming of the definition of c'; the fitness term of, for instance, the elastic net objective function does differentiate between maps A and D.

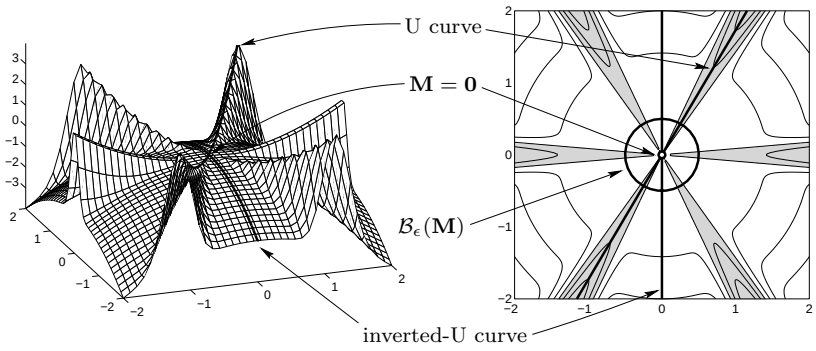


Figure 3: A function with a stationary point at the origin ($\mathbf{M} = \mathbf{0}$) that is a saddle point: (left) surface plot; (right), contour plot. The equation of the function in polar coordinates is $f(r, \theta) = r^2(\sin^{2m}(N\theta - \alpha) - \frac{1}{2})$ with $m = 10$, $N = 3$, and $\alpha = \frac{\pi}{2}$. In the contour plot, the shaded areas correspond to $f > 0$ and the white areas to $f < 0$. Any straight line in the (r, θ) plane that passes through the origin is associated with either a U curve, along which the function increases away from the origin (if inside the shaded areas); an inverted-U curve, along which the function decreases away from the origin (if inside the white areas); or a horizontal line, along which the function is constant (if on the boundary). For the particular function shown, inverted-U curves are much more abundant than U curves, and so perturbations of the point $\mathbf{M} = \mathbf{0}$ inside a small ball $\mathcal{B}_\epsilon(\mathbf{M})$ typically result in a lower value of f.

is often very close to 0 and geometrically corresponds to a direction along which f is nearly flat. This is a perennial problem in multivariate optimization (e.g., backpropagation training of a multilayer perceptron), where it can be difficult to tell whether the optimization algorithm converged to an optimum or got stuck in a saddle point or even in a nonstationary point. Bentler and Tanaka (1983) report a remarkable example from the factor analysis literature (in a space of merely 36 dimensions).

But the problem we really have is even harder, because we want to see whether a function $\mathcal{C}$ has an optimum along the manifold $\mathcal{R}(\mathbf{M})$, not along the whole space. This means that we cannot even obtain $D - 1$ linearly independent directions along which to perturb[6] the map ($D - 1$ being the dimension of $\mathcal{R}(\mathbf{M})$), as shown in section 3.2, and therefore we cannot even compute the gradient tangential to $\mathcal{R}(\mathbf{M})$. Consequently, we are not merely unable to determine whether the map is a maximum inside $\mathcal{R}(\mathbf{M})$; we cannot even determine whether the map is at a stationary point inside $\mathcal{R}(\mathbf{M})$.

[6] Assuming true small perturbations, not the ones that Swindale et al. (2000) used, as discussed in section 3.3.

3.2 Such Perturbations May Indeed Alter $\mathcal{R}$. Swindale et al. (2000) argue that the perturbations they tried should not affect the continuity measure, however this may actually be defined. The justification is that presumably such a measure would depend on the Euclidean distances between stimulus values, and these distances are preserved by the class of rigid motions. However, Swindale et al. (2000) applied rigid motions to the individual maps separately,[7] and this does alter the geometric relationships between the individual maps which are observed in biological maps. For example, the stripes of ocular dominance and orientation maps are known to intersect at approximately right angles, and the singularities of the orientation map are known to lie generally at the centers of the ocular dominance stripes (Bartfeld & Grinvald, 1992; Obermayer & Blasdel, 1993; Hübener, Shoham, Grinvald, & Bonhoeffer, 1997); and (although still awaiting experimental replication) the orientation discontinuities seem to be matched with retinotopic discontinuities (Das & Gilbert, 1997). If the individual maps are independently rotated or translated, these relationships are altered (or completely broken, if the perturbations are large). Such alterations of the individual map interrelations are likely to have an effect on the cortical wiring constraints and therefore on the value of $\mathcal{R}$. Consequently, the fact that coverage uniformity generally decreased becomes hard to interpret, since it could be accompanied by an increase or a decrease in $\mathcal{R}$.

3.3 Such Perturbations Are Not Local. So far we have used the term *perturbation* in its usual sense of small or elementary perturbation, whose amount is vanishingly small. For example, a small translation of the map $\mathbf{M}$ in the direction of a vector $\mathbf{N}$ could be defined as $\mathbf{M}' = \mathbf{M} + \epsilon\mathbf{N}$, or $\boldsymbol{\mu}'_{ij} = \boldsymbol{\mu}_{ij} + \epsilon\boldsymbol{\nu}_{ij}\ \forall (i, j) \in \mathcal{C}$, where $\varepsilon > 0$ is very small; similarly, if every $\boldsymbol{\mu}_{ij}$ is perturbed by a different small amount, then we would have a rubber-sheet perturbation.

However, the perturbations that Swindale et al. (2000) used are not small. To see this, note that their perturbations are actually *permutations*. Consider, for example, perturbing the orientation map while keeping the other maps fixed, and assume for notational convenience that the cortex origin is at the center of the rectangular region that Swindale et al. examined:

- Horizontal shift of k pixels: θ_{ij} takes the value of $\theta_{i+k,j}$.
- Horizontal flip: θ_{ij} is swapped with $\theta_{-i,j}$.
- 180 degree rotation: θ_{ij} is swapped with $\theta_{-i,-j}$.

[7] If the individual maps were transformed jointly by a rigid motion, we would gain no information (assuming that the cortex is homogeneous and isotropic).

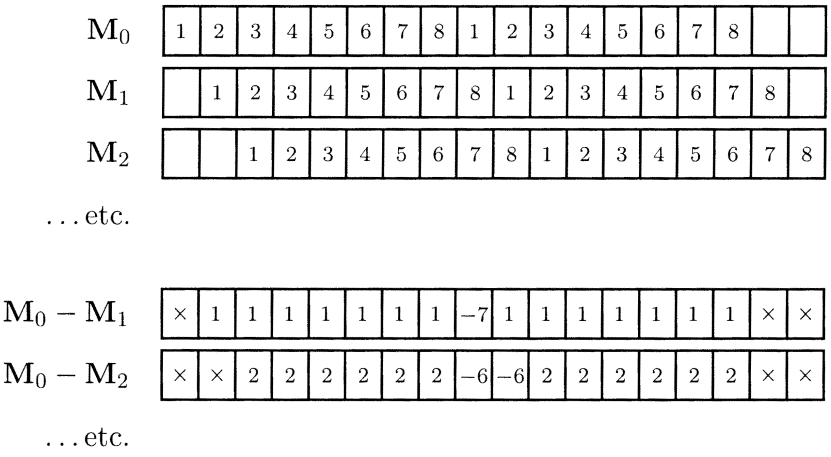


Figure 4: Shifts of a one-dimensional bit-mapped cortex with a discontinuity.

Thus, the new value for θ_{ij} will often be very different from the original one. It could be argued that for shifts of a small amount (e.g., $k = 1$ pixel or less), the value of $\theta_{i+k,j}$ will be very similar to that of θ_{ij}, but this rests on an assumption of continuity of orientation that does not hold generally (e.g., at singularities or fractures), and the same would happen with other maps. This argument holds no matter how finely one discretizes the cortex, since the discontinuities do not go away as the pixel size goes to zero. The argument would also apply to small-angle individual map rotations, though Swindale et al. (2000) considered only 180 degree rotations. Figure 4 illustrates the idea. It shows a one-dimensional bit-mapped cortex $\mathbf{M}_0$ where each pixel i contains an orientation value θ_i that varies continuously except at a single point; it then shows maps $\mathbf{M}_1, \mathbf{M}_2, \ldots$ shifted by increasing amounts of $1, 2, \ldots$ pixels and the respective perturbations $\mathbf{M}_0 - \mathbf{M}_1, \mathbf{M}_0 - \mathbf{M}_2, \ldots$ of the original map. It can be seen that the effect of both the continuous region and the discontinuity is accumulative with the shift size, consistent with the U-curves reported by Swindale et al., and if the pixel size is very small, the 1 values in $\mathbf{M}_0 - \mathbf{M}_1$ will be very small but the -7 will remain, and similarly for other shifts, so that the perturbation remains large.

In summary, the "perturbations" of Swindale et al. (2000) are not small perturbations, but permutations that often result in very large perturbations, therefore being nonlocal. That is, the perturbed map is not in the immediate vicinity of the original one but in a faraway region of map space, and therefore comparison of the coverage values of both maps is hard to interpret. Put another way, it makes sense to add 1 degree to a given θ_{ij} and see how that affects $\mathcal{C}$, but not to add unknown, potentially arbitrarily large amounts to all θ_{ij}'s. Besides, if the original (observed) map was indeed at a local optimum of $\mathcal{C}$, given the multiplicity of local optima in map space

mentioned earlier, it is to be expected that the perturbed map would then be near a completely different local optimum.

4 Conclusion

The general principle that cortical maps are wired in a way that achieves uniform coverage while also minimizing cortical wiring is important for understanding cortical map structure. The abstract implementation of such a principle in cortical map models based on dimensionality reduction replicates most of the characteristics of such maps. The evidence that Swindale et al. (2000) presented is certainly consistent with the optimization hypothesis, but it does not add significant support for it. Being based on trial and error of a subset of possible perturbations, it might disprove the hypothesis that empirical maps maximize (a certain measure of) coverage uniformity, but it cannot prove it. The lack of an appropriate definition of economy of cortical wiring prevents us from finding perturbations that leave it unchanged and ultimately prevents a quantitative assessment of the general principle stated above. What could be more easily tested is whether empirical maps are stationary points of the coverage uniformity by itself (irrespective of any connectivity constraint), but it would be surprising if this was so.

Since most of the perturbations that Swindale et al. (2000) tried worsened coverage uniformity (the more so the larger the perturbation, for the horizontal shifts), it may be argued that this cannot be due to chance. In section 3, we showed that in high dimensions, this intuition about chance is misleading and that continuity is also being altered by those manipulations. In addition, the example shown in Figure 2 demonstrates that it is also possible to observe a systematic decrease in coverage uniformity for shifts when the map is not optimal. Visual cortex has an orderly columnar structure. That perturbations such as shifts should disturb that structure, and probably worsen coverage uniformity, should not come as a surprise. That *any* perturbation should worsen coverage uniformity is a far stronger statement that would be very difficult to confirm empirically. The work of Swindale et al. (2000) provides useful evidence that coverage is fairly uniform across visual cortex, but does not prove that it is as uniform as possible.

Optimality is a very important principle in biology, and there are many examples where biological systems have been proven to achieve the best performance possible given the relevant physical constraints (e.g., Bialek, 1987). The approach in such cases is generally to calculate from first principles what optimal performance would be and then show that biology achieves this performance. This is analogous to the standard methodology in the visual cortical map field of hypothesizing an objective function (though on rather less certain grounds than direct physical constraints), calculating the maps that follow from this function, and comparing them with real maps. An alternative method, practical when the problem is discrete and the number of possible states is relatively small, is to calculate the value

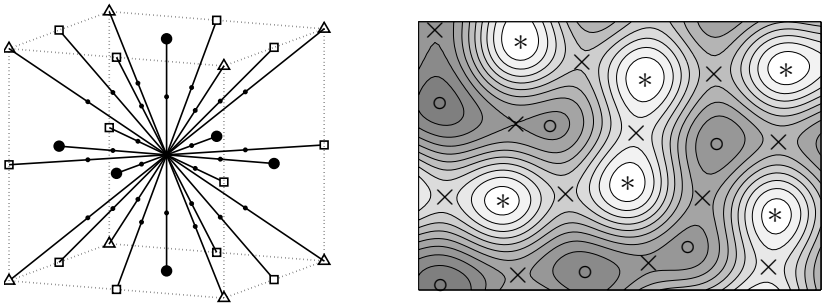


Figure 5: (Left) Setup for the 3D case of the proof of the appendix.The nearest neighbors of the central point (at the origin) of first order lie at the diagonals of length 1 (•), of second order at the diagonals of length $\sqrt{2}$ (□), and of third order at the diagonals of length $\sqrt{3}$ (△); they correspond to the 6 faces, 12 edges, and 8 vertices, respectively, of a cube of side 2 centered on the origin (dotted line). The midpoints of the diagonals are marked with small dots. (Right) Contour plot of a function of two variables with many maxima. The maxima are marked with $*$, minima with $\circ$, and saddle points with $\times$. Observe the abundance of saddle points compared to that of maxima or minima. This is because when going from one maximum to another maximum, or from one minimum to another minimum, we cross a saddle point. In higher dimensions, saddle points become even more abundant, possibly exponentially so.

of an objective function for all states and show that the biological state represents the optimum of this function (e.g., Cherniak, 1995). However, in the case discussed by Swindale et al. (2000) the number of states is infinite, and we argue that a numerical perturbation approach cannot yield significant insight into optimality.

Appendix: Abundance of Saddle Points and Optima in High Dimensions

Consider a linear superposition of localized spherical functions (each function falls away quickly, e.g., at a distance $\frac{1}{2}$) each centered at the knots of a D-dimensional cubic array. That is, we place one such function at every position $(x_1, x_2, \ldots, x_D)$ where each x_d is integer, for $d = 1, \ldots, D$. Call *whole-knot* every such position. By symmetry, we will have a maximum at every whole-knot and, over a large region, the same number of minima. In the midpoint of the segment joining any two maxima (minima), there must be either a saddle point or a minimum (maximum). These midpoints are located at positions $(x_1, x_2, \ldots, x_D)$ where at least one x_d is an integer plus $\frac{1}{2}$ and correspond to the midpoint of diagonals of lengths $\sqrt{2}, \sqrt{3}, \ldots, \sqrt{D}$ that link a maximum with its respective nearest-neighboring maxima (see Figure 5, left). Call such midpoints *half-knots*. Over a hypercubic region of side

N in each axis, we have $(2N)^D$ half-knots and whole-knots (saddle points, minima and maxima) and N^D whole-knots (maxima). Thus, it contains N^D maxima, N^D minima, and $(2N)^D - 2N^D$ saddles. Therefore, the ratio maxima:minima:saddles is $1{:}1{:}2^D - 2$, and there are $\mathcal{O}(2^D)$ saddle points per maximum or minimum. The expression is also valid for $D = 1$ (where there exist no saddle points). Figure 5 (left) shows the proof setup for $D = 3$. The groups of nearest-neighboring maxima of a maximum knot are at distances 1 (•), $\sqrt{2}$ (□), and $\sqrt{3}$ (△) and correspond to the 6 faces, 12 edges, and 8 vertices, respectively, of a cube of side 2 centered on the knot. At the midpoint of every such diagonal, there is either a saddle point or a minimum.

In less crystalline arrangements, some maxima, minima, and saddle points will coalesce, but if the function under consideration has many maxima uniformly scattered over its domain, we would expect the ratio to remain approximately correct. Figure 5 (right) shows the landscape of a bivariate function with many maxima.

Hence, saddle points are typically much more numerous than maxima and minima in high dimensions.

Acknowledgments

We thank Peter Dayan and Graeme Mitchison for helpful discussions and comments on this article, and particularly Nick Swindale for his willingness to discuss with us the issues it raises.

References

Bartfeld, E., & Grinvald, A. (1992). Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proc. Natl. Acad. Sci. USA, 89*(24):11905–11909.

Bentler, P. M., & Tanaka, J. S. (1983). Problems with EM algorithms for ML factor analysis. *Psychometrika, 48*(2), 247–251.

Bialek, W. (1987). Physical limits to sensation and perception. *Annu. Rev. Biophys. Biophys. Chem., 16*, 455–478.

Cherniak, C. (1995). Neural component placement. *Trends Neurosci., 18*(12), 522–527.

Das, A. (2000). Optimizing coverage in the cortex. *Nat. Neurosci., 3*(8), 750–752.

Das, A., & Gilbert, C. D. (1997). Distortions of visuotopic map match orientation singularities in primary visual cortex. *Nature, 387*(6633), 594–598.

Durbin, R., & Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature, 343*(6259), 644–647.

Durbin, R., Szeliski, R., & Yuille, A. (1989). An analysis of the elastic net approach to the traveling salesman problem. *Neural Computation, 1*(3), 348–358.

Durbin, R., & Willshaw, D. (1987). An analogue approach to the traveling salesman problem using an elastic net method. *Nature, 326* (6114), 689–691.

Erwin, E., Obermayer, K., & Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: A critical comparison. *Neural Computation, 7*(3), 425–468.

Goodhill, G. J., & Willshaw, D. J. (1990). Application of the elastic net algorithm to the formation of ocular dominance stripes. *Network: Computation in Neural Systems, 1*(1), 41–59.

Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of the macaque monkey visual cortex. *Proc. R. Soc. Lond. B, 198*(1130), 1–59.

Hübener, M., Shoham, D., Grinvald, A., & Bonhoeffer, T. (1997). Spatial relationships among three columnar systems in cat area 17. *J. Neurosci., 17*(23), 9270–9284.

Nocedal, J., & Wright, S. J. (1999). *Numerical optimization*. New York: Springer-Verlag.

Obermayer, K., & Blasdel, G. G. (1993). Geometry of orientation and ocular dominance columns in monkey striate cortex. *J. Neurosci., 13*(10), 4114–4129.

Swindale, N. V. (1991). Coverage and the design of striate cortex. *Biol. Cybern., 65*(6), 415–424.

Swindale, N. V. (1996). The development of topography in the visual cortex: A review of models. *Network: Computation in Neural Systems, 7*(2), 161–247.

Swindale, N. V., Shoham, D., Grinvald, A., Bonhoeffer, T., & Hübener, M. (2000). Visual cortex maps are optimised for uniform coverage. *Nat. Neurosci., 3*(8), 822–826.

Received April 24, 2001; accepted October 29, 2001.