Conjectures regarding the nonlinear geometry of visual neurons

James R. Golden[a], Kedarnath P. Vilankar[a], Michael C.K. Wu[b,c], David J. Field[a]

[a] *Department of Psychology, Cornell University, Ithaca, NY, USA*
[b] *Biophysics Graduate Group, University of California, Berkeley, CA, USA*
[c] *Lithium Technologies Inc., San Francisco, CA, USA*

Abstract

From the earliest stages of sensory processing, neurons show inherent non-linearities: the response to a complex stimulus is not a sum of the responses to a set of constituent basis stimuli. These non-linearities come in a number of forms and have been explained in terms of a number of functional goals. The family of spatial non-linearities have included interactions that occur both within and outside of the classical receptive field. They include, saturation, cross orientation inhibition, contrast normalization, end-stopping and a variety of non-classical effects. In addition, neurons show a number of facilitatory and invariance related effects such as those exhibited by complex cells (integration across position). Here, we describe an approach that attempts to explain many of the non-linearities under a single geometric framework. In line with Zetzche and colleagues(e.g., Zetzsche et al. (1999)) we propose that many of the principal non-linearities can be described by a geometry where the neural response space has a simple curvature. In this paper, we focus on the geometry that produces both increased selectivity (curving outward) and increased tolerance (curving inward). We demonstrate that overcomplete sparse coding with both low-dimensional synthetic data and high-dimensional natural scene data can re-

Email addresses: jrg265@cornell.edu (James R. Golden), kpv9@cornell.edu (Kedarnath P. Vilankar), michael.wu@lithium.com (Michael C.K. Wu), djf3@cornell.edu (David J. Field)
URL: http://redwood.psych.cornell.edu/ (David J. Field)

sult in curvature that is responsible for a variety of different known non-classical effects including end-stopping and gain control. We believe that this approach provides a more fundamental explanation of these non-linearities and does not require that one postulate a variety of explanations (e.g., that gain must be controlled or the ends of lines must be detected). In its standard form, sparse coding does not however, produce invariance/tolerance represented by inward curvature. We speculate on some of the requirements needed to produce such curvature.

Keywords: `Natural scenes`, Visual cortex, Neural networks, Deep learning

1. Introduction

Visual neurons have now been extensively mapped out with a wide range of stimuli ranging from basis sets like spots and gratings to complex stimuli ranging from white noise to natural scenes (e.g., Hubel & Wiesel (1962); Blakemore & Campbell (1969); Vinje & Gallant (2000)). The resulting responses have been shown to exhibit both increasing selectivity and increasing tolerance as one moves higher through the visual pathway (Rust & DiCarlo (2010)).

There is currently a wide variety of different descriptions of the non-linearities in primary visual cortex and they are typically presumed to serve a variety of functional goals. The various non-linearities include gain control (e.g., Schwartz & Simoncelli (2001); Ohzawa et al. (1985); Heeger (1992)), end-stopping (Hubel & Wiesel (1962); Yu & Levi (1997)), cross orientation inhibition (e.g., Morrone et al. (1982)), and a range of non-classical effects Zhu & Rozell (2013). In line with Zetzsche et al. (1999) and Zetzsche & Rohrbein (2001), we attempt to describe these different non-linearities within a single geometric framework.

By linear we mean that the response to a composite pattern is a linear sum of the responses to the basis stimuli that compose that pattern.

$$R\left(\sum_i x_i\right) = \sum_i R\left(x_i\right) \tag{1}$$

For a linear system, it is a simple process to predict the response to any complex stimulus since the response to the complex stimulus is simply the sum of the

responses to the bases that comprise the complex stimulus. To the extent that a neuron is linear, it can be treated as a vector directed into the state space of images (sometimes referred to as image space). The array of neurons selective to different image features can then be represented as an array of vectors spanning the image state space (e.g., Field (1994); Zetzsche & Nuding (2005)). The true array of neurons in V1 is neither orthogonal nor linear. However, this simplification has provided important insights into why these neurons map out the visual world as they do. One popular view is that they provide an efficient code for natural scenes. By pointing in the direction where natural scenes have the highest probability density and insuring that the space (out to some frequency limit) is spanned by those vectors, a sparse and relatively independent representation of natural scenes can be achieved (Field (1987, 1994)). Neural networks which attempt to maximize the sparseness or independence of the neural responses have been shown to result in receptive fields with properties similar to those of simple cells in the cat and primate visual systems (Olshausen & Field (1997); Bell & Sejnowski (1997)).

The approach has also met with some success in describing a number of nonlinearities in visual neurons. Overcomplete sparse coding networks have been demonstrated to generate nonlinear responses similar to behavior described as end-stopping and gain control (e.g., Hoyer (2003); Olshausen & Field (2005); Lee et al. (2006); Zhu & Rozell (2013)). Various attempts have also been made to describe invariant-like responses such as those found with complex cells in V1 and most higher level visual neurons (e.g., Hyvärinen & Hoyer (2001); EinhaÈuser et al. (2002); Berkes & Wiskott (2005); Rifai et al. (2011); Tsai & Cox (2015)).

However, in this paper we are not attempting to provide a new neural network that reproduces all the known properties of neurons in the visual pathway. Rather, we are attempting to put these various properties into a single unifying framework that allows us to account for the known non-linearities of visual neurons. We will demonstrate how a known network (i.e., sparse coding) behaves with respect to this approach and make some general proposals regarding what

higher visual areas are trying to achieve. The approach has a number of similarities to results and theories of DiCarlo and colleagues (e.g., DiCarlo & Cox (2007); Rust & DiCarlo (2010); DiCarlo et al. (2012)). These papers refer to the flattening of object manifolds achieved by the representation in higher visual areas. We will return to these issues in Section 5, but overall we argue that the curvature described in the following sections provides a means for unwrapping these manifolds.

In this paper, we provide a framework for discussing the non-linearities found in neurons in the visual pathway and apply the approach to aspects of sparse coding. We take a geometric approach and use some of the important ideas proposed by Zetzsche and colleagues (Zetzsche et al. (1999); Zetzsche & Rohrbein (2001); Zetzsche & Nuding (2005)). Like Zetzsche, we will focus on the inherent curvature in the iso-response surfaces of neurons. We will also argue that the curvature of these response spaces provides key insights into these early non-linearities and suggest there is an important relationship between overcomplete coding and this curvature.

However, we will make a number of new arguments.

1. We will argue that the sparse coding network of Olshausen & Field (1996) produces a form of curvature (curving outward, which we will later define as exo-origin curvature) that will generate many of the non-linearities found in V1. That is, this curvature results when these networks search for an efficient representation using an overcomplete code.
2. We will demonstrate that the curvature in these networks is determined by the angles between neighboring neurons and as well as the cost function and the degree of overcompleteness used in the sparse coding network.
3. Since the sparse coding network does not explicitly attempt to curve the iso-response contours (or explicitly generate the non-linearties), we argue that the geometric approach can provide a deeper insight into why such techniques generate the kinds of non-linearities found in biological systems.
4. We argue that one useful goal of this general class of curvature is to pro-

duce a representation that we call critically sampled overcomplete where an n-dimensional space is represented by $k > n$ neurons but no more than n neurons will respond to any stimulus. We show how this can be achieved in low dimensions and speculate how it might be achieved in higher dimensions.

5. We will also show that both tolerance/invariance and hyper-selectivity can fit into this framework by considering both inward curvature (which we will later define as endo-origin curvature) as well as outward curvature (exo-origin cuvature).
6. We will argue that the selectivity and tolerance of higher level neurons are represented by a combination of these curvatures.

It is important to also note that the literature exploring non-linearities in visual neurons is quite large. We could not possibly address all the papers showing the various non-linearites that have been documented. However, consider the following analogy. Imagine a 2D map of the earth (e.g., a Mercator projection) and consider the optimal travel path between any two points. On a north-south path (e.g., New York to Peru) the path is straight. One does not need a non-linear equation to describe this path. On the other hand, a trip from New York to Paris would require a significant non-linearity to achieve the optimal path. One could make a list of all the non-linearities found between any pair of cities. However, if one recognizes that all the cities lie on a sphere, then it is possible to account for all of those non-linearities within a single geometric framework. Our argument is that the current approach to understanding non-linearities in visual neurons is equivalent to making a list of non-linearities calculated from each experiment and providing a unique cause to explain each curve. We believe the better approach is to reconsider the intrinsic geometry of the neural response space and show how that compares with image state space. Although this geometry will be much more complex than a simple sphere, we believe a wide variety of non-linearites will collapse into a single overarching geometric explanation.

2. Geometry of neural response surfaces

We begin this section by repeating some of the ideas from Zetzsche et al. (1999) on several basic non-linearities. Like we suggest here, they proposed that many of the early non-linearities can be represented by simple curvatures in the iso-response surfaces. Zetzsche et al. (1999) proposed that the curvature should be considered in terms of a transformation from Cartesian coordinates to a polar representation. Both of our approaches will be considering transformations in the response space relative to the stimulus space. In this work, however, we will be focusing on the sparse coding network of Olshausen & Field (1996) and the forms of curvature produced by such a network when it attempts to find an efficient representation when the number of neurons is overcomplete. We will begin with a similar discussion to that of Zetzsche et al. (1999) of simple neural non-linearities and describe how a simple curvature produces many of the non-linearities found in the early visual system (e.g., gain control, end stopping, etc.). We will extend these ideas to descriptions to include both curvature away from the origin and towards the origin to account for both selectivity and invariance.

Mostly for simplicity, a wide range of psychophysical and modeling studies treat a visual neuron as a simple linear system. If the neuron is linear, it can be represented as a vector in the state space of all possible images. V1 is then modeled as an array of neurons (e.g., wavelet, Gabor-like, etc.) that spans the image state space. If the neuron is linear, then it is possible to capture the behavior of the neuron using a 2-dimensional image of its receptive field where intensity at each position of the receptive field represents the magnitude of response at that position. Figure 1a shows a representation of a linear neuron in a simple 2-dimensional (e.g., 2-pixel) state space where the neuron can produce both a positive and negative response. Figure 1a also plots the iso-response contours for this simple neuron represented by lines orthogonal to the vector. For such a simple system, the magnitude of the response to any stimulus is simply the projection of that stimulus vector onto the vector representing the neuron's receptive field. We emphasize this simple point, because once we warp

the space, this linear projection in stimulus space will no longer hold.

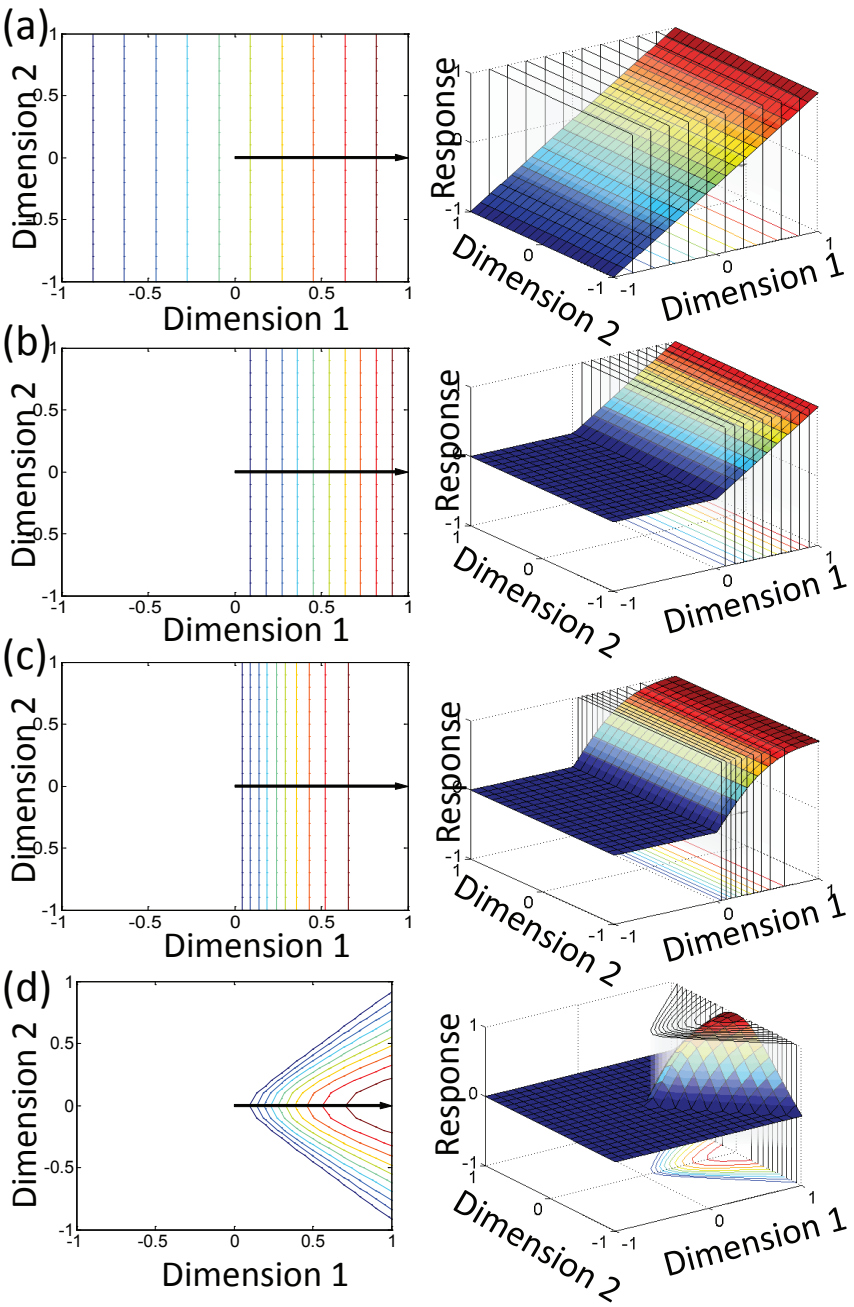

Figure 1: Please refer to the next page for the figure description.

Figure 1 *(previous page)*: The figure shows four basic types of neural response geometry in a 2-dimensional state space. On the left the neuron is represented as a vector ($[1, 0]$) in the 2-dimensional state space. Each colored line (orthogonal or curved) to the vector is an iso-response contour which represents a set of stimuli (e.g., 2-pixel images) in the image state space for which the given neuron responds with the same magnitude. The plots on the right show the response surface where the Z-axis represents response magnitude of the neuron. a) Response geometry of a linear neuron. b) Response geometry of a thresholded non-linear neuron. c) Response geometry of a compressive non-linear neuron. and d) Response geometry of a warping non-linear neuron.

For any n-dimensional image state space, the iso-response surface of a linear neuron is represented by an $n-1$ dimensional plane orthogonal to the vector. So for a three dimensional space, the iso-response surface is a two dimensional plane. For a 100 pixel image the response surface is a 99-dimensional hyper-plane which is orthogonal to the vector representing the neuron (i.e., its receptive field). Any linear operation, such as the sum of two vectors, does not alter the flatness of the iso-response surface.

2.1. Planar non-linearities (often called output or point wise non-linearities)

The first family of non-linearities that we want to consider are those with planar properties described above (i.e., the iso-response regions of the neuron are defined by a plane). These are generally described as output non-linearities or sometimes point-wise non-linearities. They can be modeled by applying a simple non-linearity to the output of the neuron. However, as we will see, even a simple output non-linearity like a threshold fundamentally alters the nature of the geometric representation.

2.1.1. Threshold non-linearities

This first form of non-linearity is the result of a simple response threshold that allows the neuron to produce only a positive response. As shown in Figure 1b, we can assume that the neuron gives no response until the stimulus intensity exceeds zero. This threshold non-linearity certainly has some biological plausi-

bility since a neuron can (in general) only produce a positive response. In this simple model, once the response exceeds zero, then the response of the neuron is linear. Although this captures a key feature of V1 neurons, it fails to account for the family of non-linearities found in the early visual system.

2.1.2. Compressive and accelerating non-linearities

Now consider what a simple form of non-linearity does to the response manifolds (where a manifold is defined by the family of iso-response surfaces). If we apply a compressive non-linearity on the output, then the iso-response surfaces are still planes. However, the distribution of the response planes representing different magnitudes will change depending on the particular form of the compression. We might consider a simple non-linearity like that shown in Figure 1c. For an output non-linearity like this, the iso-response surfaces will remain planar, but the spacing of those planes is no longer uniform. We classify this as a planar non-linearity.

A non-linearity such as this produces a compression of the spacing between iso-response planes where the neuron's contrast response is steep. However, the iso-response surfaces are still planes. A number of theories have been proposed regarding the function of this compressive response non-linearity but it is generally argued that it is either a response to the limited dynamic range of the neuron, or it is related to the contrast distributions found in natural scenes and serves to provide an efficient representation of that contrast distribution (e.g., Laughlin (1981); Brady & Field (2000); Tadmor & Tolhurst (2000)).

2.1.3. Warping non-linearites: curvature in the response surfaces

It has been noted by a number of researchers (e.g., Albrecht & Geisler (1991); Heeger (1992); Tolhurst & Heeger (1997)) that the planar non-linearities described above are insufficient to account for the behavior of visual neurons. Even the simple cells in V1 of cat and primate show significant non-linearities not described by a simple output non-linearity. Non-linear effects described as contrast normalization, cross orientation inhibition, end-stopping and other

effects outside the classical receptive field are known to require other types of non-linearity. In the section below we describe a simple warping of the iso-response planes resulting from the process of handling an over-complete basis set. We note that the resulting non-linearities describe many of the non-linear behaviors shown in cortical visual neurons.

Similar to Zetzsche et al. (1999), Figure 1d shows an example of a curvature in the response manifolds that can generate more complex non-linearities. In the following sections, we describe a variety of simple curves in the iso-response contours that allow a neuron to become highly selective or highly tolerant (or both). Figure 2 shows examples of these simple curvatures. We define two general curvature classes. Figure 2a shows examples where the iso-response contours curve away from the origin. We define these curves as having "exo-origin curvature". Figure 2b shows examples where the iso-response contours curve towards the origin. We define these as having "endo-origin curvature".

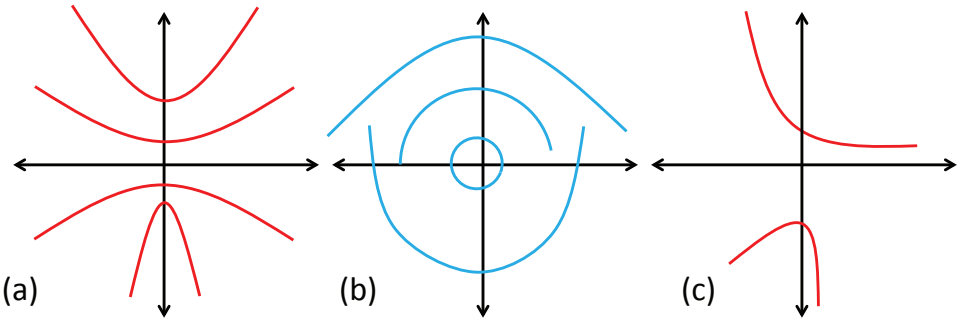

Figure 2: Examples of proposed curvatures in the iso-response contours. a) Shows examples of exo-origin curvature (curved away from the origin) and b) shows examples of endo-origin curvature (curved towards the origin). See text for details. c) Shows examples of non-symmetric iso-response contour with exo-origin curvature. Later (e.g., Figure 14d) we will discuss how the asymmetry can help produce an efficient representation with an unevenly distributed vector array.

More formally, we consider the family of iso-response contours that are curved (i.e., there exist a line segment connecting two points on the curve that does not coincide with the curve) and where the neuron's response is monoton-

ically increasing. Endo-origin curvature is defined as a curve where there exists at least two points on the the curve such that the line segment connecting the points passes through the origin. Similarly, exo-origin curvature is defined as a curve where there exists no two points on the curve such that the line segment connecting the points passes through the origin. This definition does not apply to all possible curves but applies to most of the curves that we describe in this paper.

As noted by Zetzsche et al. (1999), neurons with iso-response contours that bend away from origin (exo-origin) produces a family of nonlinearities like those found in V1. Consider what is implied by this curvature in the example shown in Figure 3. 'A' is the point representing a stimulus that elicits a firing rate of 8 spikes/second. Stimulus 'B' does not evoke any response from the neuron. However, adding stimulus 'B' (or '-B') to stimulus 'A' will reduce the neurons firing rate to 2 spikes/second. Thus, exo-origin curvature can result in non-classical receptive field responses.

This is the general form of the non-linearity described by end-stopping, and gain control (Zetzsche et al. (1999)). We argue that this exo-origin curvature describes a general family of non-linearities that we define as 'hyperselective' (which allows a neuron to respond to fewer stimuli within the image state space than a linear neuron). For this family, a stimulus which elicits no response on its own, can reduce the firing of a neuron if it is stimulated to respond with another stimulus. For example, with end stopping, an oriented bar ('A') within the classical receptive field produces a response. A bar ('B') outside of the classical receptive field produces no response. But when bar 'A' is presented in conjunction with to bar 'B', there is a reduced response.

This is also the form of curvature implied by models of cross-orientation inhibition (e.g., Bishop et al. (1973)) and gain control (Heeger (1992)). Figure 4a shows an example of data from (Albrecht et al. (2003); Albrecht & Hamilton (1982); Albrecht et al. (2002)). The neuron is presented with a variety of stimuli (gratings in this case). Some gratings produce a high response while other gratings produce lower responses. However, the contrast at which the responses

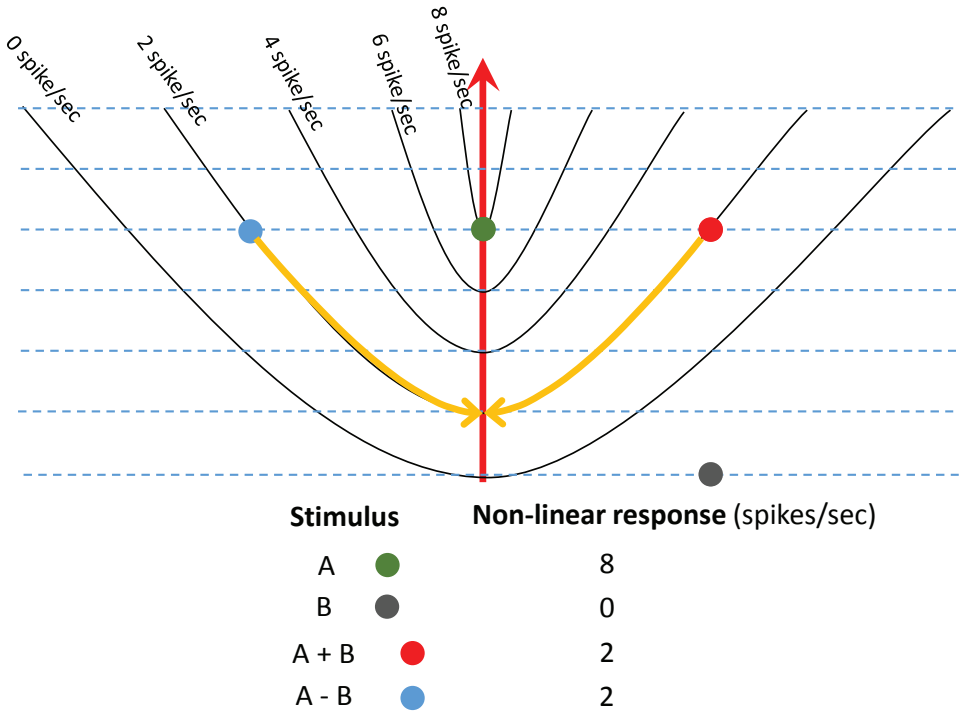

Figure 3: Exo-origin curvature and its relation to non-classical effects like end-stopping, cross-orientation inhibition, etc. In this example Stimulus 'A' represents the most effective stimulus for the neuron. For the magnitude shown, stimulus 'A' elicits $8 spikes/sec$ in the neuron. For instance, this could represent a bar presented in the center of the neuron's receptive field at its preferred orientation. Stimulus 'B' represents a stimulus that produces no response in the neuron. For example this could represent a bar presented outside the classical receptive field. Although 'B' (or '–B') produces no response on its own, when stimulus 'B' is combined with stimulus 'A', the neurons response will be reduced. Both end-stopping and cross-orientation inhibition are examples of this general form of non-linearity.

saturate are roughly the same. The solid lines in Figure 4a show the fit (from Albrecht et al. (2003)) using the Naka-Rushton (see Equation 2b) where the r_{half} and V_{max} are held constant. The stimuli used for this experiment are from an orthogonal basis (sinusoidal gratings). An orthogonal basis cannot provide information regarding the responses to stimuli between the basis vectors (i.e., it cannot see the precise form of curvature within the sub-regions of state space between the basis vectors). However, if we assume that the neuron saturates at

the same contrast for all stimuli between these orthogonal bases, then we can generate an example of response surface as shown in Figure 4b . If we consider an array of stimuli shown as the rays in Figure 4b we get the neurons responses to these stimuli as shown in Figure 4c and d . The curvature shown in these plots is a simple geometric interpretation of a system that saturates at the same stimulus magnitude (as opposed to the same response magnitude). To be more explicit Figure 4b,c and d are generated by Equation 2a.

$$resp = f_{NR}(r, n, r_{half}, V_{max}) \times \exp \frac{-\theta^2}{2 \times \sigma^2} \tag{2a}$$

$$f_{NR}(r, n, r_{half}, V_{max}) = \frac{V_{max} \times r^n}{r^n + {r_{half}}^n} \tag{2b}$$

where f_{NR} is the Naka Rushton equation with four parameters, V_{max} is the saturating value, r_{half} is the half saturation level, n is the exponent (for Figure 4b,c and d n is set to 4) and r is the contrast or radial distance of a stimulus from the origin. $resp$ is the gain-controlled response, modeled as f_{NR} multiplied by a radial Gaussian with a 0 degree mean and 30 degrees of σ. θ is the polar angle of a stimulus in a $2D$ state space.

It should be noted that not all neurons in the Albrecht & Hamilton (1982) data show this precise behavior. About 70% showed clear saturation and were well described by the Naka Rushton equation (see Equation 2b). Of these, there still remained some variation in the half saturation point. Nonetheless, the data do suggest that the majority of neurons show this general type of non-linearity. However, by mapping with only an orthonormal basis, we can not see the precise form of curvature of the response manifold. For example, knowing only three points that lie on a $2D$ surface within a three-dimensional space is insufficient to define the curvature of that surface. More than three points are required to determine whether the $2D$ surface is curved. Similarly, to determine the curvature of a real neuron's response manifold, we must have a relatively dense mapping of responses like that shown in Figure 4b, c, and d. We are not aware of physiological data that provide this detailed mapping.

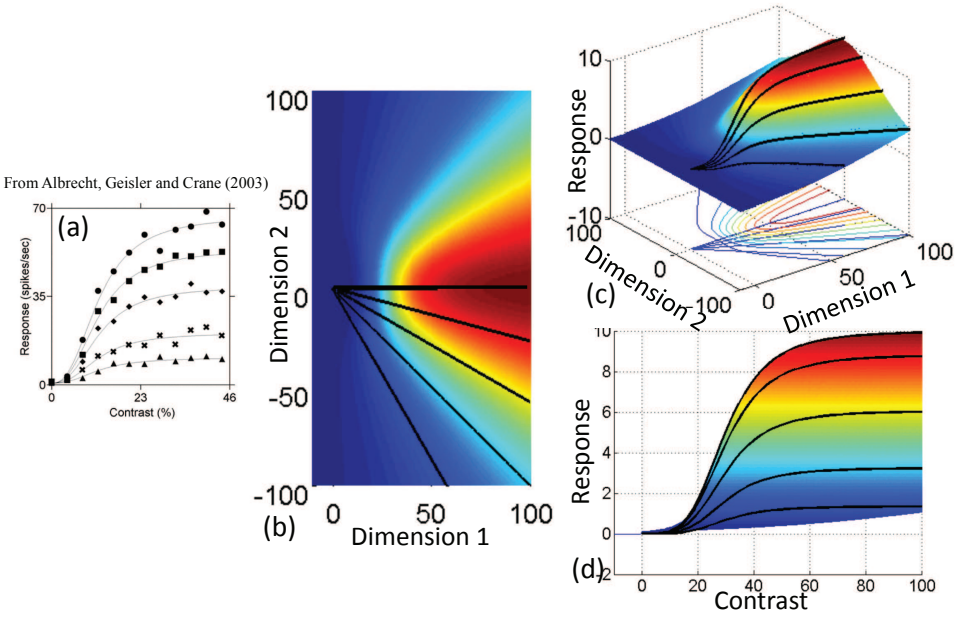

Figure 4: The geometry of gain control. a) shows the classical results derived from Albrecht et al. (2003) that show a neuron that saturates at different response magnitudes for different stimuli (gratings of different spatial frequencies) but saturates at roughly the same stimulus magnitude (contrast). These data show the response to an orthonormal basis (sinusoids). If we make the assumption that the neuron saturates at the same contrast for all stimuli between these orthonormal bases, we can generate a response manifold. b) shows a heat map representing this response manifold, if we assume that the Naka Rushton equation (Equation 2b) describes the contrast response for all stimuli. A ray extending from the origin represents a particular stimulus of varying contrast. c) shows a side view of this response surface along with the iso-response contours of the neuron and d) shows the contrast response generated with this response surface, where contrast is defined as the distance of a point from the origin. As one can see with this response surface the neuron saturates at the same stimulus contrast. This is not intended to be an accurate model of the curvature of the original neuron (probed with a high-dimensional orthogonal basis). Rather the intention is to demonstrate the relation between the geometry and the contrast response. These response surfaces b),c) and d) are generated using Equation 2a where the selectivity is defined by the σ. For this example, we have assumed the curvature is symmetric around the primary axis of the neuron. As we will discuss, this assumption does not necessarily hold.

The main point of the proceeding discussion is to suggest that what may appear to be very different non-linearities (e.g., end stopping and gain control),

each with their own assumed goals are, from this perspective, the same general non-linearity. This suggests that a general theory for these non-linearities may be possible without appealing to a variety of different neuronal functions. Later, we will demonstrate that this curvature also has important implications for how the array of visual neurons tiles the stimulus space. The exo-origin curvature limits the population of stimuli that will stimulate the neuron. Therefore, a full tiling of the space requires an overcomplete code (a point noted by Zetzsche & Rohrbein (2001)). In our interpretation, this curvature provides a means of isolating the sparse causes of the data and providing an efficient means of mapping the space.

It is worth noting that there are a variety of possible exo-origin curvatures. In Figure 5, we show five examples of different curvatures and the effects they have on the responses to different stimuli. In Figure 5 the sub-figures on the left show a side-view of the neural response surface and the sub-figures on the right show the top-view of the iso-response contours for the corresponding response surface on left. On each plot, a single black ray extending from the origin represents a stimulus of varying contrast. The stimuli radially distant from the origin have comparatively higher contrast than the stimuli near the origin. Different rays represent different stimuli (different combinations of Dimension 1 and Dimension 2). In Figure 5a, we have a simple geometric representation of data like that shown in Figure 4d where the response to all stimuli saturates at the same contrast. We have simply re-plotted the fit to the contrast response data as shown Figure 4a. To help distinguish these different models, we have normalized the average curvature of each model to have roughly the same magnitude(same degree of selectivity). As one can see, the full response surface of each model (shown by the family of iso-response curves portrayed on right) can have very different overall shapes.

Figure 5b shows an example of a simple version of a divisive normalization model. The standard model of divisive normalization (Heeger (1992)) involves the division of each neuron by the activities of its neighbors. If we assume that the neuron has only one orthogonal neighbor we can plot the effect this division

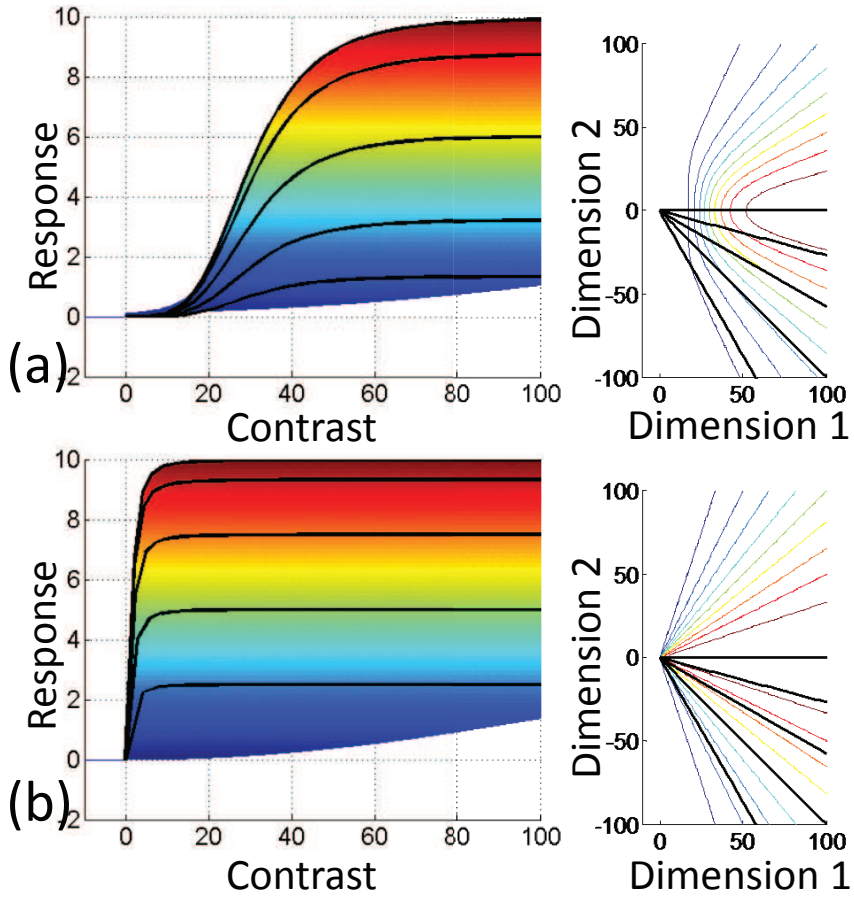

Figure 5: Five types of non-linearity and their geometric representations. On each plot, the rays extending from the origin represent stimuli of varying contrast. The stimuli radially distant from the origin have comparatively higher contrast from the stimuli near the origin. a) The curvature of a neuron that corresponds to the physiological response described as gain control. This is a simple re-plotting of the data where we assume that each response curve represents the response to an array of stimuli that are increasingly divergent from the optimal stimulus for the neuron. b) The curvature produced by a simple version of divisive normalization described in Equation 3. The figure continues on the next page.

has on the response profile. The output is defined as:

$$resp = \frac{r_1}{\frac{r_1+r_2}{2} + 1} \tag{3}$$

where r_1 and r_2 are the squared linear responses of two orthogonal neurons.

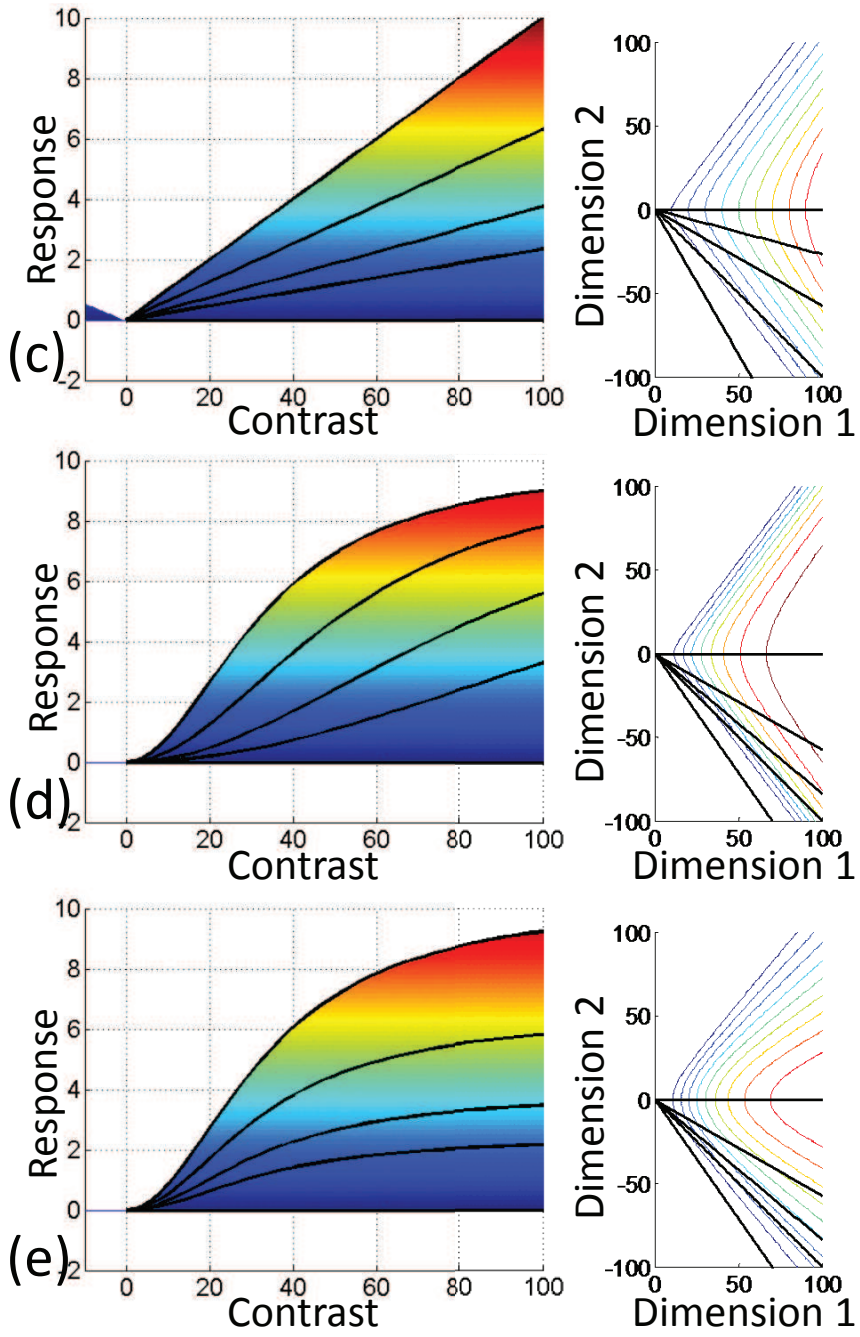

Figure 5 *(previous page)*: c) The curvature produced by the warping non-linearity described in Equation 4. d) The curvature produced by the warping non-linearity described in Equation 4 with output compressive non-linearity described in Equation 5. e) The curvature produced by the warping non-linearity described in Equation 4 with input compressive non-linearity described in Equation 6. In all of the examples provided, we have normalized the magnitude of curvature to be roughly the same. The magnitude of curvature in each model is a free parameter (except for the standard divisive normalization model shown in b)) and each can therefore produce a variety of magnitudes. However, each model will produce a different curve family which we have emphasized by normalizing the magnitude. In the supplemental materials (Figure S8), we provide examples of each model when the magnitude of curvature is varied.

For Figure 5c-e the curved response surfaces are generated by the fan equation:

$$a_i = f_{Fan}(c, \theta) = c \times \cos(n(f_i, f_j)\theta)$$
$$n(f_i, f_j) = \frac{\pi/2}{\arccos(\frac{\langle f_i, f_j \rangle}{\|f_i\| \|f_j\|})} \tag{4}$$

where c is the distance of a stimulus from the origin (i.e. the stimulus contrast), θ is the angle between a stimulus and the neuron, a_i is the response magnitude of a neuron i, n determines the curvature and f_i and f_j are the vectors representing the two neurons. n is a function of the angle between neighboring neurons. When $n = 1$ the iso-response contours are flat (e.g., linear); when $n > 1$ the neuron has iso-response contours with exo-origin curvature and when $n < 1$ the neuron has iso-response contours with endo-origin curvature. In Section 5 we will return to this issue of how the non-linearity relates to the neighboring neurons.

The left part of Figure 5c shows the response surface of a neuron defined by Equation 4. One can note from the figure, as the contrast increases, the response also increases linearly and never saturates. In order to saturate the response a compressive non-linearity could be applied either to the output of the neuron or to the input to the neuron. Left plot on Figure 5d shows the response surface of a neuron which has compressive non-linearity applied to the response computed using Equation 4. This response is computed as:

$$resp_i = f_{NR}(f_{Fan}(c, \theta), n = 2, r_{half}, V_{max}) \tag{5}$$

where f_{Fan} is the response of a neuron measured using Equation 4 and f_{NR} is the compressive non-linearity defined in Equation 2b. Similarly the compressive non-linearity could also be applied to the input to the neuron.

Figure 5e left figure shows the response surface of a neuron which has compressive non-linearity applied to its input and then the response computed using Equation 4. This response is computed as:

$$resp_i = f_{Fan}(f_{NR}(c, n = 2, r_{half}, V_{max}), \theta) \tag{6}$$

where f_{Fan} is the response of a neuron measured using Equation 4 and f_{NR} is

the compressive non-linearity defined in Equation 2b. Our goal with this model was to show how the curvature would change if the input had a saturating non-linearity (e.g., a saturating photo-receptor nonlinearity) and this was followed by our fan equation. As one can see, the behavior of this neuron has similarities to the gain control model (see Figure 5a) where the neuron begins to saturate at roughly the same stimulus contrast.

Figure 5 demonstrates one of the advantages of the geometric approach, as it allows a straightforward side by side comparison of these five types of non-linearities. For example, with this simple version of divisive inhibition, the higher magnitude iso-response curves show increasing curvature. This implies that at higher contrasts, the neuron is more selective. The particular curvature is likely to have important implications for coding the scene. However, in this paper, we are not arguing explicitly for any one of these curvatures. We should also note that both in Figure 4 and 5 we have assumed that the curvature is symmetric about the vector representing the neuron's optimal response. In Section 5.2.1 we will argue that the curvature can be asymmetric (see also Figure 14d and supplemental Figure S4) and depends on the relative angles between the neuron under consideration and each of its neighbors.

At this time, we do not believe there are sufficient physiological data to distinguish between these models. As we mentioned earlier, the use of an orthonormal basis does not allow us to examine the precise curvature of a neurons response surface. Unless one measures the responses to stimuli that are at angles less than 90 degrees apart in state space (less than orthogonal) one cannot determine the true shape of the iso-response contours. Rarely is such mapping done. This requires a much larger population of stimuli. Even when using white noise as stimuli, the angles are rarely much less than 90 degrees. In the chromatic domain where the state space is essentially 3-dimensional, it is more feasible to provide a dense mapping of the stimulus space. For example, Horwitz & Hass (2012) measured the non-linear chromatic response of V1 neurons in the three-dimensional state space defined by the cone photoreceptors. In line with the above discussion, they found evidence for both exo-origin and endo-origin curva-

ture in the iso-response surfaces as well as planar (linear) responses. We believe it is possible to provide a similar mapping of the early visual neurons (e.g., V1) but it requires a very specific set of stimuli concentrated at the curved regions of the response space. However, capturing this high-dimensional curvature in image space will require a large number of stimuli.

We will return to this issue of exo-origin and endo-origin curvature in Section 5 when we consider our ideas regarding how this curvature is used by the visual system to tile the space. In the next section we explore the curvature produced by the non-linearities found with the sparse coding network.

3. Mapping the responses of a sparse coding network

In general, the sparse coding network (Olshausen & Field (1996)) forms a reconstruction of the input data I with an overcomplete basis set Φ weighted by activations a, such that the reconstruction minimizes a sum of the squared error and the $L1$ norm of the activations over a large dataset. The parameter λ controls the sparsity of the population response. The model was first formalized using an energy equation:

$$E = [preserveinformation] + \lambda \times [sparsenessofa_i] \tag{7}$$

$$E = \frac{1}{2}|I - \phi A|^2 + \lambda \sum_i S(a_i) \tag{8}$$

The energy equation can be derived by formulating a generative model of the input data Olshausen & Field (1997). ϕ and a are learned using gradient descent on the energy function (see Equation 9 for gradient descent on a). $S(a_i)$ is the cost function which constrains the optimization to have a sparse solution. There are various forms that the cost function can take, such as $-e^{-x^2}$, $\log(1 + x^2)$ and $|x|$, which yield similar results (Olshausen & Field (1996)). These reflect different assumptions about the form of the sparse prior on the activations a in the generative model.

The network learns an overcomplete representation of the data, which means that the number of columns in the basis ϕ is greater than the dimensionality of the input data I. This was the important difference between the sparse coding network and the early versions of the Independent Components Analysis (ICA); the latter did not allow for an overcomplete representation, and produced a simple linear transform of the input data (although there are now overcomplete ICA algorithms (Hyvarinen (1999); Le et al. (2011a)). The original forms of sparse coding and ICA are quite similar when the number of basis functions is equal to the dimensionality of the input and when there is no noise (Olshausen & Field (1997)). Therefore, the critically sampled sparse coding network (where the number of basis functions is equal to the dimensionality of the input data) will also find a linear representation of the input data. The overcomplete sparse coding network, on the other hand, uses a "sparsifying" rule to learn the best reconstruction while minimizing the redundancy created by that overcomplete code.

The overcomplete representation that the network learns is nonlinear, and it has been demonstrated that the nonlinearities give rise to non-classical receptive field effects like end-stopping, cross-orientation inhibition and related non-classical surround suppression (Zhu & Rozell (2013); Lee et al. (2006)). In order to provide an intuitive understanding of the geometric form of these nonlinearities, we first explore the geometric layout of the response surfaces when the sparse coding network is applied to sparse data in a low-dimensional ($2D$) state space.

In order to visualize the iso-response contours, we examined a simplified sparse network for 2-dimensional data. We created a $2D$ sparse dataset with three causes (directions in which data points lie) and trained a sparse coding network (Olshausen & Field (1996)) using three basis functions ($1.5\times$ overcomplete) to encode the data. We consider the case where data lies sparsely along three directions in the $2D$ state space (Figure 6a). A 2-vector description of these data will not be able to efficiently map such data. A 3-vector description as shown in the figure will align the vectors (neurons) with the data. Figures

6b and 6c show the results of the sparse coding network when two levels of λ are used. The λ parameter determines how sparse the representation is.

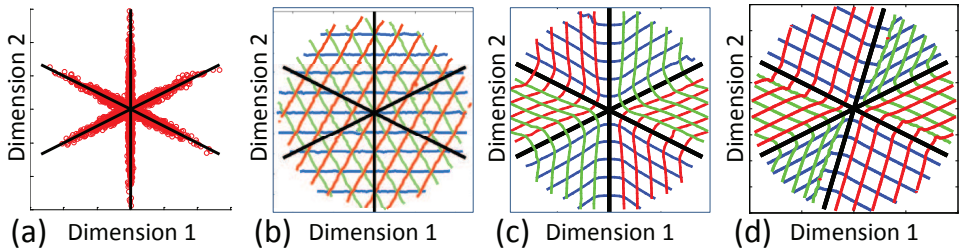

Figure 6: Results of an overcomplete sparse coding network applied to synthetic 2-dimensional sparse data. a) Scatter plot of $2D$ sparse data set where the set results from three sparse causes represented by the three axes. b) and c) Results of the sparse coding network with three basis vectors (1.5× overcomplete). The plots show the iso-response contours for each of the three neurons. b) shows the result when $\lambda = 0.01$. c) shows the result when $\lambda = 0.25$. With higher λ the network puts more emphasis on finding a solution that is sparse. The network's representation is a result of a recurrent nonlinear computation. As one can see, the iso-response contours have exo-origin curvature. This results in a representation where no more than two neurons are active for any given stimulus. Iso-response contours of each neuron are shown with different colors. d) shows the result when the causes are not symmetrically distributed. As one can see the curvature that is learned is asymmetric. However each region of the space is represented by no more than two neurons.

To evaluate the network, we determined the iso-response contours of the encoding vectors by measuring the response of the network for points uniformly distributed throughout the $2D$ state space. They are plotted for the low and high sparseness (λ) for this 1.5× overcomplete network. The iso-response contours are not curved for the low λ condition; however, they are clearly curved for the high λ condition. This response of an overcomplete network (resulting from the sparsifying function) produces the non-linear responses corresponding to the exo-origin curvature discussed in the previous section. However if the λ is near zero then the responses of neurons in the network will be nearly linear even when the network is overcomplete.

Clearly, the sparse coding network is warping the iso-response surfaces as λ is increased. The network has found an efficient set of $2D$ subspaces to describe

these data with reduced redundancy. As one can see in Figure 6c, when λ is high no point in this state space is represented by more than two neurons. That is for any stimulus, no more than two neurons are active. We will describe this scenario as "critically sampled overcomplete" when n-dimensional data are represented by k vectors ($k > n$) and no more than n vectors respond to any given stimulus. Exo-origin curvature can allow for this form of overcomplete representation. As we will see in the next section, the particular mapping and curvature depend on the cost function especially in high dimensions.

3.1. Evaluating the theory in high-dimensional natural image space

In this section, we explore the behavior of the sparse coding network for high-dimensional (8×8-pixel) natural scene data. Before we describe the results, a note should be made regarding the angles between neurons with overcomplete high-dimensional data. In Figure 6 we described the results for a $1.5X$ overcomplete data. For such $2D$ data (e.g., 2-pixel), increasing the number of vectors from 2 to 3 decreases the angle between neighboring vectors from 90 degree to 60 degrees. However, in high dimensions, increasing the number of vectors results in a much smaller reduction in the angle between neighboring neurons. For a 64 dimensional data set, an $2\times$ overcomplete representation (128 vector), the angles between neighboring neurons reduce only marginally from 90 degrees. Even if we choose a random set of 128 vectors in a 64 dimensional space, the average angle between any vector and its 64 nearest neighbor is 84 ± 0.75 degrees (the histograms are available in the supplemental material (Supplementary Figure S1). The implication of this for our approach is that even with a significantly overcomplete code the typical angle between neighbors decreases only slightly. Therefore with high-dimensional data the predicted curvature is expected to be low relative to the simple $2D$ plots we have shown in the previous section.

To explore the responses of the sparse coding network for a high dimensional data set, we applied the network to 8×8 image patches drawn from natural scenes using the standard network described by Olshausen & Field (1996). Figure 7a shows the typical output of such a network to natural scenes. To estimate

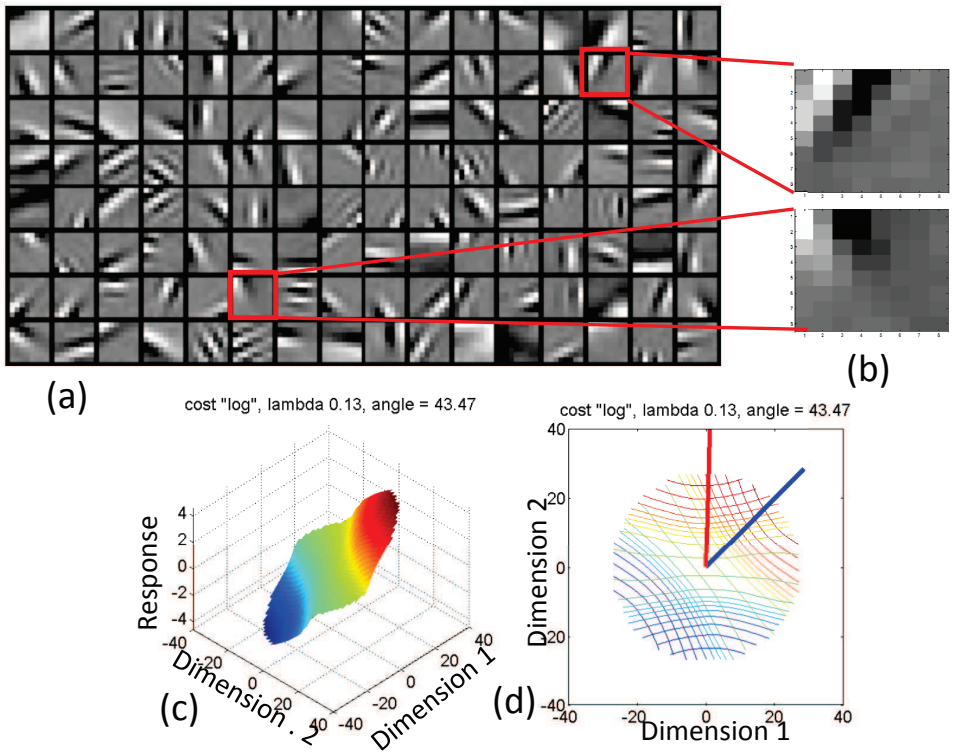

Figure 7: a) Learned basis set (128 neurons) of an ovecomplete sparse coding network in a 64 dimensional input space trained on natural scenes. b) shows an example of two receptive fields from the learned basis set that have angles less than 90 degrees (i.e., less than orthogonal). For these two neurons, the angle is 43 degrees. c) shows the response surface for one of the neuron in the $2D$ subspace defined by the two vectors (basis). d) shows the iso-response contours for both the neurons in the $2D$ subspace. The contours show clear evidence of exo-origin curvature warped around the neuron's optimal response.

the iso-response contours, we considered each of the possible pairs of neurons and measured the angle between them. The circular whitening filter used in the Olshausen & Field (1996) network reduced the dimensionality by approxi-
440 mately 40% resulting in roughly 38 independent dimensions. These data were represented by 1.7× overcomplete code producing 64 vectors for the 38 dimensions.

In the $2D$ example shown in Figure 6, the curvature that was learned is re-

lated to the angle between the neighboring vectors. The result of this curvature is to produce a representation where every stimulus is represented by no more than 2 neurons (every point in the image state space is uniquely represented by the values of only two iso-response contours). For the 64-dimensional sparse coding network, we explored the curvature by probing the 2-dimensional subspace defined by every possible pair of neurons. Figure 7 shows one example where there was significant overlap in the $2D$ receptive fields, resulting in less than 90 degree angle between the vectors that represent the receptive fields. Figure 7d shows the iso-response contours in the $2D$ subspace between these two neurons.

The basis functions were learned through training on whitened 8×8-pixel natural image patches for both critically sampled networks and for networks with several degrees of overcompleteness (1.7, 3.4 and 5.2 times overcomplete). Then, for all pairs of basis functions, a $2D$ subspace of the $64D$ image state space was determined, and the response for both basis functions was determined for data over a uniform grid in the $2D$ subspace. Each of the test data points is a point in $64D$ space and therefore represents a possible stimulus image in the $2D$ subspace between the pairs of basis functions. To map out the subspace requires probing the network with a large number of test stimuli. This is not a problem in this computational simulation. However to provide a similar mapping for real neurons, this dense probing of the space would likely be prohibitive.

It is important to keep in mind that for the sparse coding network the response of any neuron to a stimulus depends on the response of the rest of the neurons in the network. The activations a_i for each neuron are determined by performing a gradient descent on the energy function (Equation 8). The gradient step for each a_i is defined by the following differential equation:

$$\dot{a_i} = b_i - \sum_i C_{ij} a_j - \frac{\lambda}{\sigma} \dot{S}(a_i) \tag{9}$$

where $b_i = \sum_{x,y} \phi_i(x,y) I(x,y)$ and $C_{ij} = \sum_{x,y} \phi_i(x,y) \phi_j(x,y)$. In this implementation S was $\log(1 + x^2)$, as in Olshausen & Field (1996).

The response surface for the two neurons shown in Figure 7b are plotted in

Figure 7c and d. Figure 7d shows the iso-response contours for the two neurons plotted on this $2D$ subspace. As one can see the iso-response contours have exo-origin curvature with the high selectivity aligned with the direction of the vector (the neuron's optimal response). In agreement with the $2D$ example in Figure 6, this suggests the higher dimensional sparse coding network also warps the response space when the angle between the pairs of neurons is less than 90 degrees.

3.2. Measuring curvature for the high dimensional solutions

Our earlier discussions have focused on the general shapes of the neural response manifolds produced by different non-linear models. In this section, we provide a more quantitative analysis of the iso-response contours that are produced by the sparse coding network. In high dimensions, the measurement of curvature is not straightforward. There are a variety of ways to probe the 64-dimensional response space generated by a 64 dimensional sparse network. In this paper, we focus on the $2D$ subspaces defined by each pair of vectors in the code. With 64 vectors, we have a total of $n(n-1)/2$ or 2016 possible pairs of vectors. For each pair, we probe the $2D$ space defined by this pair. An example of this space is shown in Figure 7d. We can fit any given contour with a parabolic equation ($y = ax^2 + b$), where the neurons primary axis (receptive field) is normalized to coincide with the y axis. Since the iso-response contours can be asymmetric about the primary axis, we describe only the fit in the sub-region between the two vectors (where the angle between the vectors are less than or equal to 90 degrees). For example in Figure 7d, we are fitting iso-response contours only in the region between red and blue vectors. In many cases the full curvature may not be well fit by a symmetric parabola. However this approach allow us a way to compare the curvature generated by sparse coding networks with the predicted curvature from the fan equation.

Figure 8 shows the magnitude of curvature (parabolic fit parameter a) plotted as a function of the angle between the neighboring neuron, with each of the four plots showing different degrees of overcompleteness. Overall, the results

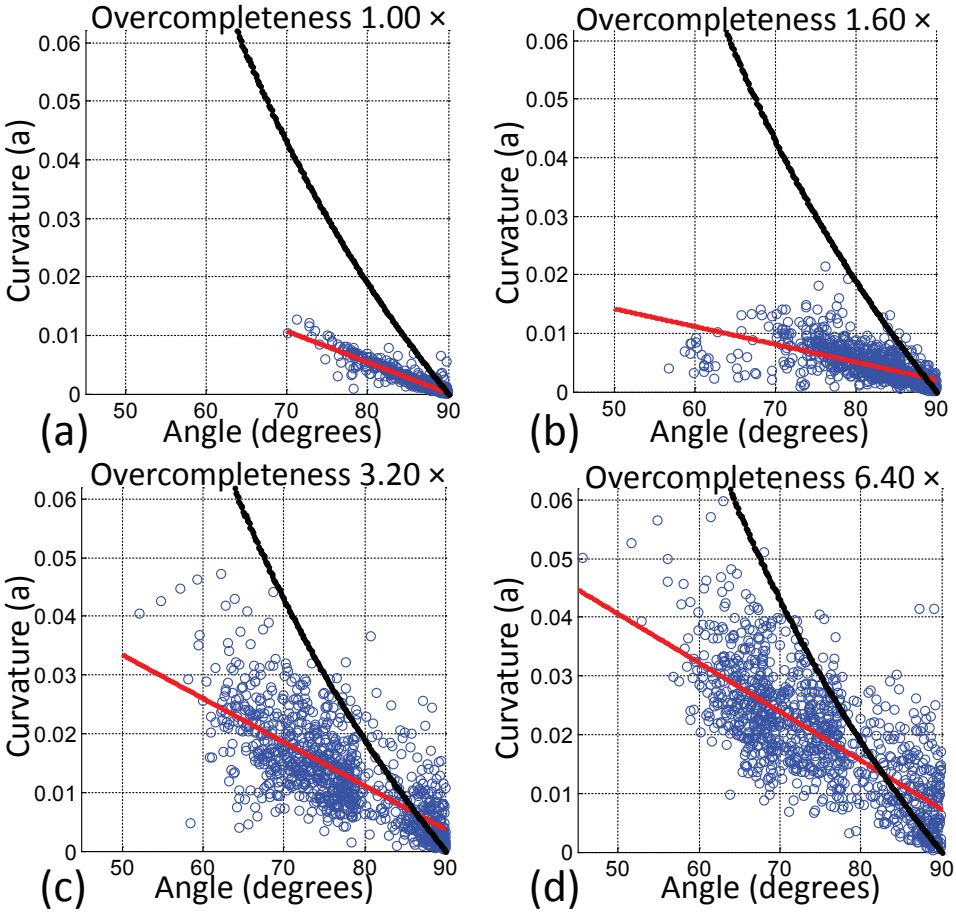

Figure 8: The four figures show the curvature of the iso-response contours for the sparse coding network when trained on natural scenes. Curvature was measured in the two-dimensional sub-regions defined as the region between each pair of learned neurons (vectors). Results are shown for four degrees of overcompleteness using a measure of parabolic parameter a (see text). Note that the curvature is at a minimum for vectors that are orthogonal (90 degrees). For angles less than 90 degrees the curvature increases (higher exo-origin curvature) with decreasing angle. As the representation becomes more overcomplete we find more neurons with a high degree of curvature. The red line shows a linear fit to the data, and the increasing slope of the line with overcompleteness of the network indicates that curvature generally increases with overcompleteness. The black lines in each of the figures shows the predicted curvature of the iso-response contours generated using the fan equation (Equation 4). As one can see the curvature with the sparse coding network is less than that predicted by the fan equation.

show that in all cases the curvature increases as the angle between neighbors decreases. The red line shows the linear fit to the data, and the black line shows the predicted curvature from the fan equation for the $2D$ sub-region. Notice that the fan equation makes the same prediction of curvature for all degrees of overcompleteness since the fan equation depends on only the angle between neighbors. In all four plots, the curvature in the sparse coding network is less than that predicted from the fan equation. We believe that this implies that the sparse coding network is not achieving the most efficient representation (i.e., more neurons may be active than are necessary). However we are still exploring how this curvature functions in high-dimensional spaces. Also one can see that with higher degree of overcompleteness (e.g., Figure 8d) there is higher relative curvature. With a higher degree of overcompleteness there is an increased probability that there are multiple neighbors less than 90 degrees away and we believe that this population of neighboring neurons can increase the curvature. We also note that there is considerable variability especially when the degree of overcompleteness is high (for Figure 8d the correlation between curvature and the angle between neurons is 0.33). With a highly overcomplete code there exist multiple ways to represent a given input (the solution is degenerate). We believe this degeneracy is responsible for a significant amount of this variability.

3.3. The effect of the cost functions on curvature

It is important to note that the precise curvatures of the neurons in the sparse coding network depend on the cost function. Figure 9 shows the iso-response contours for a neuron in the sparse coding network with two different cost functions: ($\log(1+x^2)$ and $-\exp(-x^2)$). As one can see, the cost function has a significant effect on the response curvature. With the $\log(1+x^2)$ cost function (Figure 9a) the curvature is relatively constant for the different iso-response contours. However, for the $-\exp(-x^2)$ cost function (Figure 9b) the curvature is greater for the higher iso-response contours. The behavior with this cost function is more similar to the gain-control model shown in Figure 5a. We are currently exploring the role of the cost function in determining how

the network curves the space. For the analyses in the next section we consider only the results using the standard $-\exp(-x^2)$ cost function. However, it is important to recognize that each set of results depends on the cost function. In fact, these differences demonstrate the power of using the geometric approach to understand the network properties. Körding et al. (2003) have argued the cost function produces little change in the basis functions (receptive-fields) that emerge from the sparse coding network. However, these results demonstrate that the full response surfaces may depend critically on cost functions.

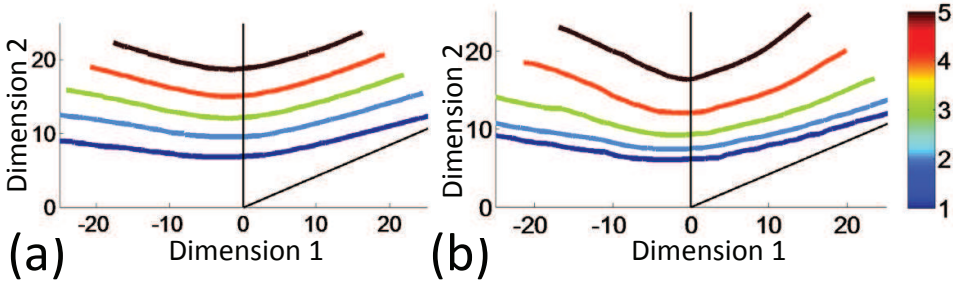

Figure 9: The iso-response contours for a neuron in the sparse coding network with two different cost functions. The cost functions are $(\log(1+x^2)$ (a) and $-\exp(-x^2))$ (b). In this example the angle between the two neurons is 60 degrees (shown by the two line segments extended from the origin). See text for details.

4. Endo-origin curvature and its relationship to tolerance and invariance

In the previous section, we showed that an overcomplete sparse coding network will produce an exo-origin curvature in its response surface which is related to the general class of hyper-selective non-linearities. In Section 5, we will provide an approach to understanding these non-linearities and make predictions regarding how these non-linearities relate to the overlap between neighboring neurons. In this section, we would first like to consider the issue of endo-origin curvature and how it relates to the tolerance/invariance of visual neurons. However, it is important to note that we do not believe that the sparse coding network provides an effective means of learning this tolerance/invariance. As we

note in the discussion (Section 6), there are number of techniques that can learn to develop tolerance (e.g., Berkes & Wiskott (2005); Karklin & Lewicki (2005)), but the sparse coding network is not one of these.

4.1. Endo-origin curvature

The exo-origin curvature described in previous sections creates a neuron that is hyper-selective to the input stimuli. The neuron will respond to a smaller set of possible stimuli than a linear neuron with a planar/linear geometry. It is similar but not identical to an AND-like process across different stimulus features. We propose that those non-classical effects that reduce the firing of the neuron (e.g., end-stopping, cross orientation inhibition, etc.) are described by a geometry with exo-origin curvature. However, these certainly do not account for all the known non-linearities of visual neurons. Neurons show a number of forms of facilitation in addition to the inhibition described above. In V1, the most common form of these is represented by neurons described as complex. These act as an OR-like process. This represents a form of invariance or tolerance where the neuron will respond to a larger proportion of images than would be expected from a linear neuron. Such as facilitatory non-linearities can be produced with a simple change in the curvature of iso-response manifolds. The endo-origin curvature is shown in Figure 10.

In macaque, complex cells are believed to represent about 3/4 of neurons in area V1 (Kagan et al. (2002)). Relative to simple cells, these neurons show a degree of positional tolerance. When presented with drifting sinusoidal gratings, these neurons shows little or no modulation. The popular energy model (Adelson & Bergen (1985)) produces this behavior by modeling the response of a complex cell as the vector sum of an even and an odd symmetric neuron. Although such models do not provide a complete account of the behavior of the complex cells in response to natural scenes (Prenger et al. (2004); Touryan et al. (2005)), the energy model has proved to be an effective first approximation.

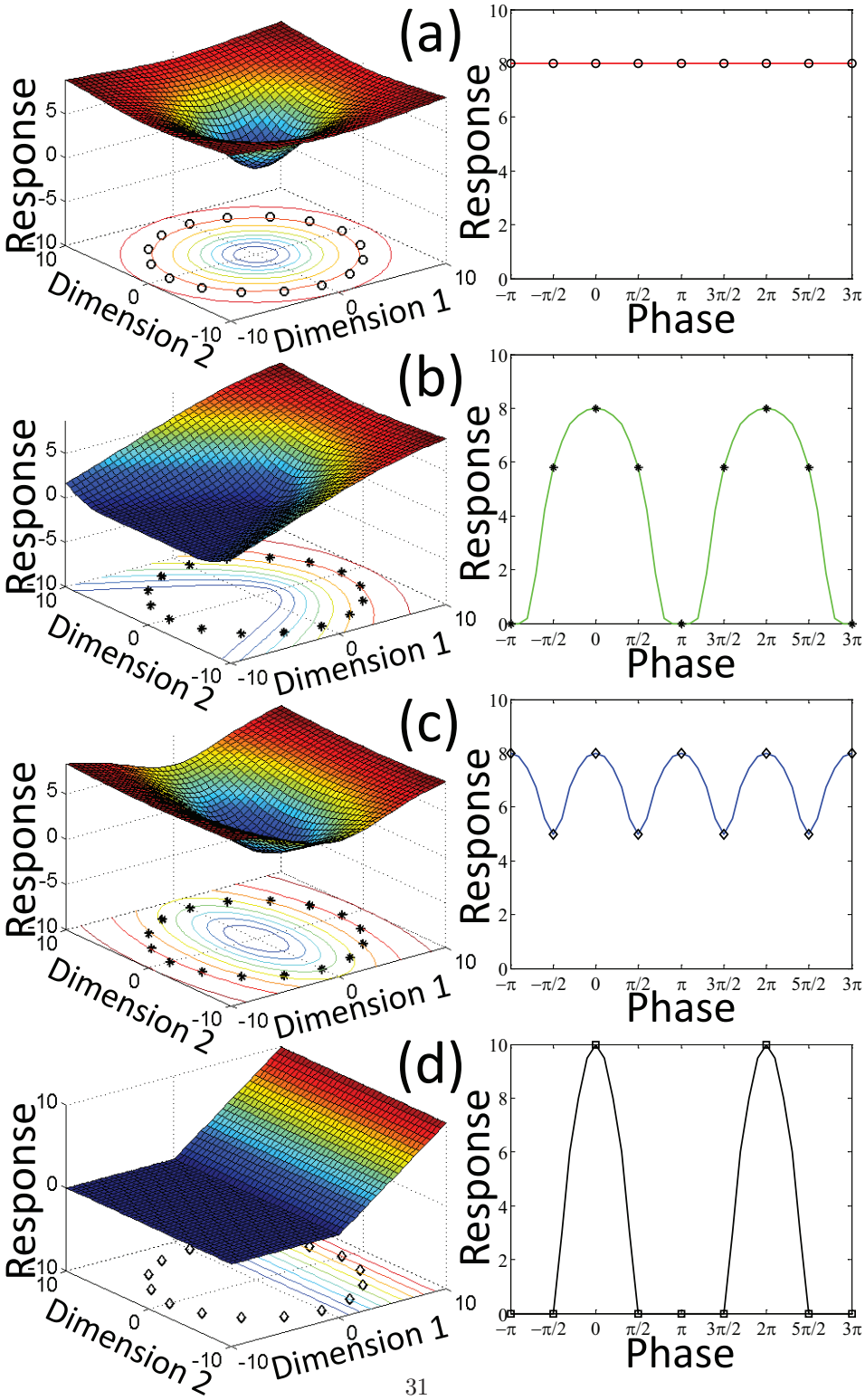

Figure 10: Please refer to the next page for the figure description.

Figure 10 *(previous page)*: Modeling a V1 neuron's response to a drifting sinusoidal grating using four models of endo-origin curvature. a) represents the classic complex cell (energy model) which has circular iso-response surface. The response to a drifting grating is flat as shown on right. d) shows the geometry of a linear simple cell (thresholded) and its half rectified response. b) and c) represent models of neurons that bridge the range between simple and complex cells. V1 neurons show a range of behavior between simple and complex (e.g., Dean & Tolhurst (1983))

Figure 10a shows a two-dimensional example of the state space of a complex cell based on the energy model. If we define the two axes as the even and odd symmetric simple cells, then the iso-response contours are circles. Drifting a grating across the receptive field will produce a constant response independent of the phase of the grating. However, this is only one version of a complex cell. Strict classification of V1 neurons into only simple: 'no tolerance' and complex: 'complete invariance to a drifting sinusoid' does not hold. When one plots the degree of modulation for drifting sinusoids, there exists an almost continuous distribution of modulations although that distribution is bimodal (e.g., Dean & Tolhurst (1983); Skottun et al. (1991)). There remains debate as to whether simple and complex neurons are distinct populations, or two ends of continuous distribution. However, our goal is not to settle this debate, or to provide a complete account of all complex cell properties. Here, we wish to show only the geometric interpretation of this variation. A circular iso-response contour describes the constant response as one varies the phase of a grating. It represents the vector sum of the underlying even and odd receptive fields in the classic energy model.

Figure 10 shows two examples of endo-origin curvature. Figure 10a shows a closed curvature where the response is invariant to the phase of the sinusoidal. Figure 10b shows endo-origin curvature that is not closed. In this case, the response of the neuron shows greater tolerance to the phase of the grating, but like many complex cells, this response will oscillate with the phase of the grating.

Of course, this two dimensional plot can show only a cross section or a pro-

jection of the high-dimensional space. However, we can generalize this geometry to higher dimensions. For a standard energy model of a complex cell, the even and odd symmetric inputs completely define the response of the neuron. Stimuli orthogonal to these two neurons will produces no response and such stimuli will not alter a response when added to stimuli that do elicit a response. Figure 11 shows a 3D visualization of this. Since the response to any stimuli along the orthogonal dimensions does not alter response of the cell, the iso-response surface will form a cylinder. The cross section of the cylinder is defined by the vectors representing the even and odd-symmetric components. In all other directions, the cell's response is independent of the stimulus magnitude and therefore flat.

Neurons that show invariance/tolerance across a variety of orthogonal inputs with the same contrast imply an endo-origin curvature. This is an OR-like behavior describing a neuron that accepts a variety of possible stimuli. For a neuron to show tolerant or invariant behavior (e.g. translation, size etc), endo-origin curvature is required.

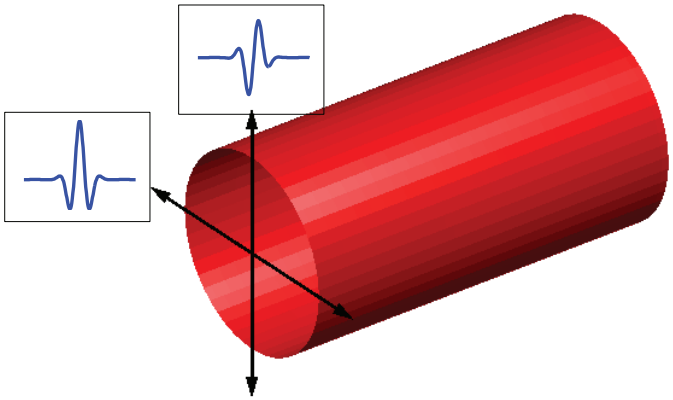

Figure 11: Geometry of a classic model of a complex cell in a 3-dimensional state space. Summing the squared responses of quadrature pair filters creates a circular iso-response contour. In all other dimensions the response is independent of the magnitude. This implies that the iso-response surface is flat in all other dimensions. The figure shows an example with one other independent dimension.

We do not want to imply that neurons can show only endo-origin or exo-origin curvature. In fact, we would argue that most neurons (especially beyond

V1) will show both tolerance/invariance and hyper-selectivity suggesting both endo-origin and exo-origin curvatures. Figure 12 shows an example of the iso-response surface of a neuron that has both endo-origin and exo-origin curvature. This describes a neuron that is invariant along two dimensions (like a complex cell) but hyper-selective in a third dimension (e.g., a complex-cell with end-stopping). In higher dimensions, this is more complex but generally follows the same logic. Neurons can be both hyperselctive in some dimensions (i.e., exo-origin curvature) and tolerant in other dimensions (i.e., endo-origin curvature). As we discuss in the next section, this approach allows us to provide a description of neurons that become more selective and more tolerant as one moves to the higher levels of the visual system.

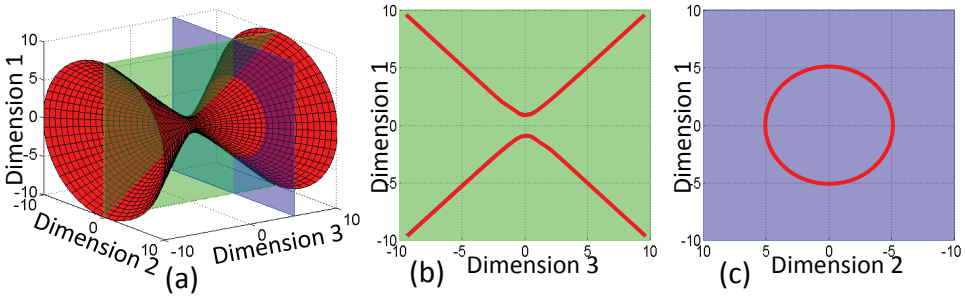

Figure 12: a)Geometry of a neuron with both hyper-selective/exo-origin curvature and invariant/endo-origin curvature b)Cross section of the iso-response surface along Dimension 1 and Dimension 3 showing exo-origin curvature (hyper-selective). 2 showing endo-origin curvature (invariant/tolerant). For example this would represent three dimensions of a complex cell with end stopping or gain control along Dimension 3

5. The proposed goal of these curvatures

In section 3, we demonstrated that an overcomplete sparse coding network will produce exo-origin curvature in response to simple sparse data as well as natural scenes. In section 4 we argued that neurons that have endo-origin curvature show tolerance/invariance to the input. In this section, we provide a more detailed hypothesis of why these curvatures are produced and discuss some

implications and predictions of this approach. Our goal here is to argue for a general approach to understanding the purpose of these non-linearities without appealing to a family of disparate sub-goals (e.g., gain-control end-stopping, etc.).

We begin our discussion with a few general comments on the statistical structure of natural scenes. Our goal in this section is not to produce an accurate rendition of the state space of images. Rather, we wish to show what we believe is the general problem that the early visual system must solve. We use simple cartoons with a basic three-dimensional state space, but we believe that these cartoons can help provide insights into the functions of non-linearities in V1 neurons.

5.1. Probability density functions and image manifolds

In the previous section, we focused on the curved manifolds representing the responses of neurons within the state space of images. Here we focus on the probability density function (PDF) of natural scenes and consider the distribution resulting from particular transformations (translation, size, etc) on the objects in those scenes. To the extent that this is a smooth probability distribution, we can refer to this distribution as an image or object manifold. Natural scenes occupy only a tiny fraction of the space of possible images (Chandler & Field (2007); Hosseini et al. (2010)). However, the distribution is far from a Gaussian cloud of points. The early work on sparse coding emphasized this non-Gaussian distribution (Field (1987, 1994)) and has focused on the star-like nature of the image manifold. Here we want to begin the address aspects of the manifold beyond the basic star shape.

In general, we are in accordance with the proposal that one of the primary goals of the ventral (V1 to IT) pathway is to perform object recognition. In line with DiCarlo & Cox (2007), we argue that the manifolds representing different objects are highly entangled when described in pixel space or receptor space. The question we address in this section is how the visual system achieves this untangling. Here we describe a very general process that has two main com-

ponents: the system must separate the object manifolds (discrimination) and respond consistently to a wide range of variations of that object (invariance or tolerance). We argue that one of the primary ways this is achieved is through the curvature in the neural response manifolds. We approach this issue using very simple geometries.

Consider the possible data sets represented in Figure 13. Each of these figures shows examples of data sets for images in $3D$ state space (e.g., 3-pixel images) that have been implied by various approaches to sensory coding (e.g., see Field (1994)). Drawings like these are fundamentally limited since they can represent only a $2D$ projection of a 3 dimensional distribution. However, such cartoons can be useful in describing the general nature of the problem that must be solved. The first data set (Figure 13a) is Gaussian. In this example, we show a data set where the pixels are correlated. All projections of these data will be Gaussian. The second data set (Figure 13b) shows a simple example of sparse data. If we think of each axis as a cause, then this state space represents three independent causes. With data sets like this, sparse coding and ICA will find the directions corresponding to the causes. In this case, each 'cause' is represented by a single direction in the image state space with some probability distribution. The third image represents an over-complete set of causes where each cause is a single direction in image state space. The number of causes is larger than the number of dimensions. This is the kind of data set we used to train the sparse coding network in the previous section (Figure 6).

Figure 13d and 13e provides examples of what we believe is the more general problem that a visual system must solve using the early non-linearities. As with Cox and Dicarlo (DiCarlo & Cox (2007)) we are arguing that object recognition in the visual system involves the untangling of a high-dimensional object manifold. However, we prefer to use the term "unwrapping" to imply that the separation and flattening of the object manifolds can be achieved with a relatively smooth and regular process. We are suggesting that the non-linearities in V1 begin this process. Figure 13d shows a single curved manifold. The particular curvature that we show is not important (we can not show a true object

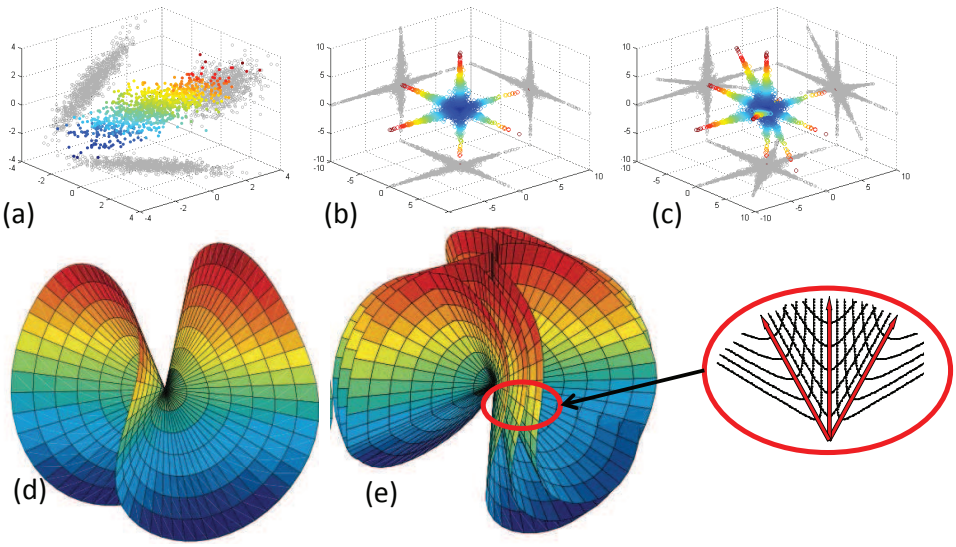

Figure 13: Five models of images in $3D$ image state space (e.g., three pixel images). a) Correlated Gaussian data b) Sparse data with the causes equal to the number of dimensions and the causes are independent. With such data a neural code with independent outputs is possible. c) Overcomplete sparse data with the causes greater than the number of dimensions. Each cause in this distribution was sampled independently, however the linear projection of these data demonstrates that no linear representation will produce independent outputs. d) and e) Two cartoons representing the general problem that we believe the visual system must solve. In d) we provide a caricature of a single object manifold in high dimensions, where the curvature of the surface represents the possible transformations on the object (e.g., translations). These are not intended to be accurate portrayals of actual object manifolds. Rather, they are intended to show only a general problem that an object recognition system must solve. In e) we show a caricature of three object manifolds representing three similar objects (e.g., three faces) undergoing the same transformations (e.g., translations). We also show our solution for separating the manifolds discussed in Section 5

manifold in just three dimensions). The argument is that a single object when transformed by position, size, orientation etc, produces a smoothly curving (differentiable everywhere except the origin) manifold in the image state space (see Rao & Ruderman (1999); Miao & Rao (2007); Wang et al. (2009) for their discussion of Lie groups). Figure 13e represents multiple objects in that $3D$ space. Consider two similar objects (e.g., two faces). When the two faces are spatially

aligned, the dot product of the two faces would show that the two vectors are pointing to similar directions in state space (i.e., they have significant dot product). As we translate each of the faces, the transformation in state space will move along similar smooth curves. We can think of the two manifolds representing the two faces as each rolled up and moving together through a constant curve in the high dimensional space as we translate the faces. We believe that if one can find a process that allows the "unrolling" of one of the face manifolds, it is likely to help unroll the other. This implies that a number of object manifolds can be unrolled with the same general algorithm and can begin at relatively early stage of visual processing (V1). Overall, we argue that the visual systems early non-linearities attempt to achieve two goals. First, they attempt to separate the manifolds by directing the neural responses at the sparse causes of the data and applying the exo-origin curvature shown on the right of Figure 13e. Second they attempt to integrate along the object manifolds using the endo-origin curvature described above (producing invariance or tolerance to the physical variations of the object). Again, we must emphasize that these simple $3D$ cartoons do not capture the rich high-dimensional structure of real objects or natural scenes. However, we will argue that this simple geometry will provide important insights into the range of non-linearities in the early visual system.

We should note that in this paper we are not proposing a technique to learn an invariant representation. A number of studies have proposed possible solutions (Rao & Ruderman (1999); Miao & Rao (2007); Wang et al. (2009); Berkes & Wiskott (2005)). Our goal here is to only note the issues that must be faced with a system that is building a more selective and tolerant representation. We must point out that the tiling of the image state space with two types of curvature is not straightforward. Below we describe one approach to tiling image space with exo-origin curvature using the fan equation (Equation 4). However, efficient tiling with neurons that have multiple curvatures is a fascinating high dimensional jigsaw puzzle.

However, there exist a number of object recognition algorithms (e.g., Serre et al. (2007)) that learn solutions that have both high selectivity and high tolerance

to object transformations. We are currently investigating these networks to determine the particular curvature that is learned by these techniques. The network of Karklin & Lewicki (2005) learns some degree of invariance from natural scenes and they propose that the network can produce complex cell-like behavior as well as more complex forms of tolerance. Our initial work suggests that the learned neurons do produce a geometry with both exo-origin and endo-origin curvature similar to that shown in Figure 12 (Golden et al. (2016); Golden (2015)).

5.2. Efficient tiling of image state space

Here we consider an approach to efficiently tile the image state space using an overcomplete set of descriptors and the curvature described above. The argument we describe below was noted by Field & Wu (2004) and discussed in Olshausen & Field (2005)). We should also note that Zetzsche et al. (1999); Zetzsche & Nuding (2005) do argue that an array of neurons with concave-like curvature do require an overcomplete code. In this section, however, we will be making a specific proposition that relates the exo-origin curvature to the angle between neurons (angle in state space) and we argue that this angle is related to the directions of the sparse structure of the data.

The basic proposal is described in Figure 14. We argue that the exo-origin curvature is primarily a means of reducing the redundancy caused by using an overcomplete code of natural scenes. Consider the number of vectors needed to represent a particular set of data with dimensionality D (e.g., an 10×10 pixel image patch would have $D = 100$). No matter what form the data take, such data never require more than D linearly independent vectors to represent it. A system where data with dimensionality D are spanned by D orthogonal vectors is described as critically sampled. Such critically sampled systems (e.g., orthonormal bases) have proven both useful and popular in the image coding community as they allow any input pattern to be represented uniquely, and the transform and its inverse can be easily computed. However, the visual system is not using a critically sampled code. In cat V1, for example, there are roughly

25 times as many output fibers as there are input fibers from the LGN, and in macaque V1, the ratio is on the order of 50 to 1. Such overcomplete codes have one potential problem in providing an efficient representation: the neurons cannot be linearly independent. In an overcomplete code with linear neurons one will typically have more neurons responding to any given stimulus than are necessary to uniquely represent the stimulus.

Figure 14 shows the mapping produced by Equation 4. In Figure 14b, we show a standard Cartesian representation with three neurons representing three directions of state space. If neighboring neurons are orthogonal, then they will have flat iso-response contours ($n = 1$ in Equation 4). For neurons that are orthogonal, there is no need for curvature. The space is well described with linear neurons or neurons with simple planar non-linearities (e.g., saturating non-linearities). However, as the angle between the neurons decreases the curvature in the iso-response surface increases. As shown in Figure, 14a and c, the iso-response surfaces are warped in proportion to this angle.

Equation 4 describes one proposed transformation which we have described as the fan equation. The curvature of the iso-response contour is determined based on the function $n(f_i, f_j)$, where f_i and f_j are the vectors representing the neurons in image state space. This particular equation describes the following transformation. If we considered Figure 14b to represent an open fan,then Figure 14c shows how the iso-response contours would curve if the fan was closing. As the primary vectors (shown in red) are moved closer together (i.e., closing the fan) the curvature increases. In this approach, the curvature in the iso-response surfaces is entirely dependent on the angle between neighboring neurons. Exo-origin curvature is produced only when the angle is less than 90 degrees ($n(f_i, f_j) > 1$). The iso-response surface between neurons i and j for a particular response a_i is represented by the fan equation. One advantage of this equation is that the total RMS response ($\sum R_i^2$) to any stimulus is unchanged by the warping of response space (i.e. the distance from the origin remains constant independent of the opening or the closing of the fan).

The problem with this approach is that although this describes the behavior

of the sparse coding network in low dimensions, the approach does not provide a good account of gain control behavior in V1 neurons without additional constraints. As was shown in Figure 5c, the fan equation does not produce response compression.

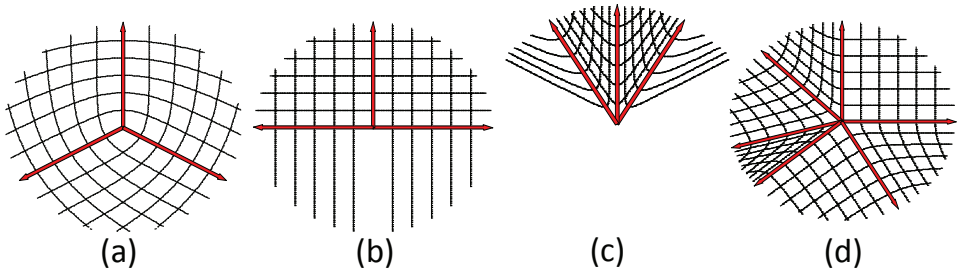

Figure 14: A family of curvatures produced by Equation 4. a)Shows Iso-response contours with endo-origin curvature resulting from vectors more than 90 degrees apart. b)Shows flat iso-response contours resulting from orthogonal vectors. c)Shows iso-response contours with exo-origin curvature resulting from vectors less than 90 degrees apart. d) Shows the proposed solution resulting from the six non-negative vectors distributed asymmetrically. As one can see, the different $2D$ subspaces can have different amounts of curvature resulting in asymmetric iso-response contours.

Consider a case where we have sparse data (like that originally shown in Figure 6a). These data are two-dimensional but have 6 non-negative causes that are independent with each cause having a 1 in 6 probability of having a non-zero response. Figure 15a shows the curvature in the iso-response contour learned by the sparse coding network (replotted here from Figure 6c). Figure 15b shows the predicted warping as described by Equation 4 for the six vectors aligned to the causes. As one can see, for this $2D$ case, the theoretical prediction from the fan equation provides a good fit from the results from the $2D$ sparse coding network. For this $2D$ state space, the response surface (the family of iso-response contours for a given neuron) predicted by Equation 4 accounts for 97% of the variance produced by the sparse coding network. Figure 15c shows the results when a compressive non-linearity (defined in Equation 6) is applied to the input before the fan equation is applied. The curvature now allows gain control behavior where the neuron saturates at the same response magnitude.

As we have shown in Figure 9 (and Figure S2), in high dimensions, the curvature that is produced with the sparse coding network depends on the cost function. The curvature behaves more like 15b when the cost function is $\log(1+x^2)$ (see Figure 9a), but behaves more like the Figure 15c when the cost function is $-\exp(-x^2)$ (see Figure 9b). We are currently exploring the role of the cost function in more detail. However, both of the approaches in Figure 15b and c provide methods for projecting the data into six $2D$ subregions. Each reduces the redundancy created by using an overcomplete code.

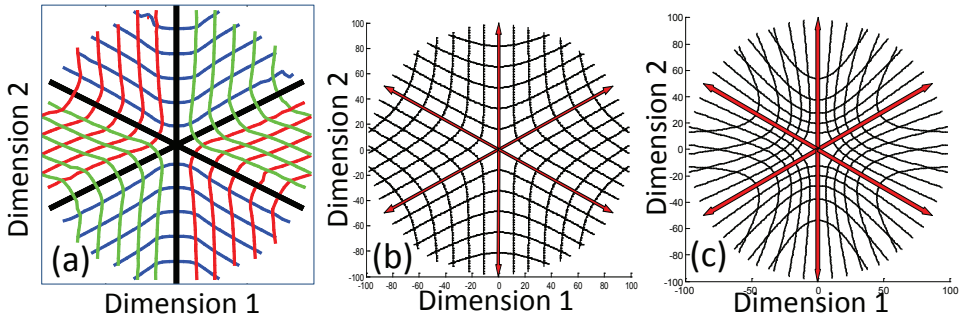

Figure 15: A comparison of the sparse coding results with two equations of warping curvature. a) Results from the sparse coding network replotted from Figure 6c. b) Iso-response contours for each neuron (vector), where the neuron's response was determined using Equation 4 and c) Iso-response contours for each neuron (vector), where each neuron's response was determined using Equation 4 where input to the neuron had compressive non-linearity defined in Equation 6. As one can see in c) the addition of a compressive input non-linearity results in increased curvature at higher contrast. As noted in Figure 4 this increased curvature can produce contrast response curves that are more closely matched to the results from contrast gain experiments.

By warping the space around the neuron in this way, we create a situation where for any stimulus, no more than 2 neurons are active at any time. This is the case even though the representation is overcomplete. Overall, for any data with image dimensionality N (e.g. $N = 100$ for 100 pixel images), one would theoretically never need more than N vectors to describe any image. However, with linear vectors, and an overcomplete basis, there will be more than N non-zero projections. The solutions we show in Figure 15 provide approaches of

assuring that any data point can be represented within a 2-dimensional subregion. We believe that it is possible to generalize these equations to allow any point in an N dimensional state space to be represented by no more than N vectors. We are currently working on ways to generalize Equation 4 into high dimensions. However with an overcomplete code, the solution is not trivial since there are multiple ways to tessellate the high-dimensional state space.

5.2.1. Asymmetric solutions

As we have previously noted, the sparse coding solutions do not necessarily produce symmetric curvature. The fan equation will also not produce a symmetric solution when the neighboring vectors (i.e., neighboring neurons) are asymmetrically distributed around the primary response vector of the neuron (the angles to the neighboring neurons are not equal). Figure 14d shows a solution described by Equation 4 where there are six unevenly distributed non-negative causes. Because the angles between the neighboring neurons are not equal, the six different $2D$ sub-regions can have different amounts of curvature resulting in asymmetric iso-response contours. We have found that the asymmetry that is produced by sparse coding is quite similar to that predicted by the fan equation. However in high-dimensions, the solution is much more complex. We are currently investigating how the family of neighboring neurons alters the full curvature of the neuron's response manifold.

Although we believe the fan equation makes reasonable predictions in many cases, we are not trying to make a strong argument for a particular form of curvature. Each will have its strengths and weaknesses. In sparse coding, the particular form of curvature depends on the cost function, and the degree of overcompleteness. Furthermore, the true curvature of V1 neurons has yet to be determined. Overall however, we believe that the exo-origin curvature increases as the angle between the neighbors decreases.

6. Discussion

In this paper, we have argued that many of the early non-linearities are described by simple curvatures of the iso-response manifold of neurons. Following from the work of Zetzsche et al. (1999) and Zetzsche & Nuding (2005), we have argued that non-linearities like end-stopping, cross orientation inhibition and related non-classical effects can be described in terms of an exo-origin curvature. Gain control and divisive normalization are also examples of this exo-origin curvature although as we show in Figure 5, the details of each model have important implications for the precise form of the curvature. Nevertheless, the fact that these different non-linearities have the same general geometric property invites the question of whether they serve the same general computational goal.

In the previous sections we have argued that the exo-origin curvature allows an efficient sparse representation of data using an overcomplete code. Our fan equation (Equation 4) describes a method that guarantees that no more than 2 vectors are active with an 2 dimensional dataset even though the code is overcomplete (total number of vectors$>$ 2). We are currently working to generalize this equation to higher-dimensions, where no more than N vectors have a non-zero response. We believe that this approach provides an account of a number of early non-linearities without the need to invoke different functional goals for the different instantiations of the implied curvature. For example, consider the function of the curvature implied by gain control models. We argue that this curvature is useful for an overcomplete representation whether or not the neurons have limits on their dynamic range. It provides an efficient means of tiling the space with an overcomplete code. Therefore, this general gain control behavior is useful even when there is no need to control the gain of neurons. Furthermore, as Zetzsche has noted, if neurons are using this concave-like curvature then they are responding to smaller regions of the image state space than a similar linear neuron. This implies that even if one argued that the goal was to control gain, a complete representation with such neurons would require an over-complete code. We believe the sparse coding approach is a more

succinct approach for describing the non-linearities. As shown in Figures 14 and 15, the curvature depends directly on the angle between the neighboring neurons. Our prediction is that when two neurons are orthogonal there is no curvature in the subregion between the neurons. As the angle is decreased, the curvature is increased.

In Section 3, we demonstrated that the original sparse coding network of Olshausen & Field (1996) produces this theoretical relationship between the curvature and the angle between neighboring neurons. The network searches for directions in state space that maximize the sparseness of the population of neurons. However, the sparse coding network also uses an overcomplete code where there are more neurons than are needed to critically sample the image state space. We demonstrated that for simple low dimensional spaces, it is possible to visualize this curvature directly. For higher dimensional solutions, we described the results when the sparse coding network is applied to natural scenes. In an overcomplete code, for any given neuron, there will be a number of other neurons that are not orthogonal (less than 90 degrees apart). We can explore the curvature in the subregion between any pair of neurons. As shown in Figure 8 and as predicted, the smaller the angle between neurons, the greater the curvature.

In Figure 5 we showed 5 non-linear models that produce curvature. One of the advantages of this geometric approach is that it allows us to easily compare the impacts of the different non-linearities on the response geometry. At this time, however, we are not attempting to promote one model over another for V1 neurons. We do not think we can infer the precise curvature of a neuron's response manifold using results generated from an orthogonal basis (e.g., sine-waves). What is needed is a large family of stimuli near the region in image state space that produces the most effective response. This will require probing this area with a large number of stimuli that span the regions between any pair (or n-tuple) of neurons. White-noise stimuli are likely to be ineffective since the probability is low that one will probe the appropriate region of state space with a sufficient number of stimuli. The use of natural scenes may increase the

probability of probing the appropriate regions of the image state space, but the probability is still quite low.

We believe, however, that these ideas are certainly testable experimentally. We are currently exploring how one might optimize that family of stimuli needed to acquire information about curvature. In this paper, we have mostly focused on the non-linearities in V1. However, we are confident that these geometric ideas will be useful in describing non-linearities in V2 and V4. Both selectivity and tolerance have increased in these areas (e.g., Rust & DiCarlo (2010)) and we argue that this implies a higher degree of curvature in these areas.

We have argued that both selectivity and tolerance/invariance are important for understanding early non-linearities. In this paper most of the discussion has been directed at selective/exo-origin curvature. As we noted, we do not believe the sparse coding network is an effective means of learning tolerance. However there have been various techniques that allow networks to learn tolerance (endo-origin curvature). The "OR-like" operations found in algorithms like slow feature analysis (Berkes & Wiskott (2005)) or the recent work of Karklin & Lewicki (2005) discussed earlier describe possible methods for learning this endo-origin curvature. Other approaches that attempt to perform object recognition while tolerant to the typical transformations of that object (e.g., Serre et al. (2007); Le et al. (2011b); Yamins et al. (2014)), will likely have both endo-origin and exo-origin curvature. We are currently probing the units developed by these networks to measure the sorts of curvature they produce (Golden et al. (2016); Golden (2015)). We are also currently developing our own algorithm that learns both endo-origin and exo-origin curvature from natural scenes using simulated eye-movement data. These models bring up a number of issues related to the problems of tiling the image state space with neurons that have multiple curved dimensions.

We should also emphasize that we have so far ignored a range of parameters relevant to V1 neurons. We have not discussed temporal sensitivity, binocular selectivity or chromatic selectivity. Temporal sensitivity is certainly an extremely important issue that must be addressed in future papers on this topic.

We have treated the neurons in the visual pathway as using a rate code with static non-linearities. This is certainly a gross oversimplification. There are many ways that temporal selectivity might interact with the non-linearities described here. The curvature may well have a temporal component in biological neurons. If we believe the curvatures are implemented biologically through lateral and feedback connections, then the curvature may change over time. We might find that the curvature is higher a few milliseconds after the initial response. This would result in more information in the spike train than provided by a simple rate code.

In a similar vein, we believe this approach may provide insights into visual adaptation. Again, if we think of the lateral and feedback connections as the cause of curvature then manipulations that alter the activity of surrounding neurons is likely to alter the curvature. There remains a question as to whether the geometric approach will prove useful in describing the full spatiotemporal response of visual neurons. Overall, however, we believe it has the potential to unify a variety of non-linear behaviors into a single framework. We also believe this approach can be applied to different sensory systems (e.g., auditory or somatosensory neurons) and provides a framework for understanding how networks learn efficient and invariant representations of the world.

7. Summary

In this paper, we have provided a geometric approach to visual non-linearities that provides a means of exploring high selectivity and high tolerance in visual neurons and neural networks. Along the lines of Zetzsche et al. (1999), we suggest that a number of early non-linearities (e.g., end-stopping, contrast gain control) can be placed within a single framework where the iso-response surfaces have a exo-origin curvature relative to linear/planar neurons. We have further argued that the sparse coding network of Olshausen & Field (1996) produces this curvature as a means of producing an efficient representation with an overcomplete code. We have found that the sparse coding network produces

curvature similar to that of a folding fan. In our approach, we make the testable prediction that the exo-origin curvature is directly related to the angle between neighboring neurons in the state space. However, there remains a question of the curvature of actual V1 neurons. Measuring curvature in real neurons requires a much more dense distribution of stimuli than what is typically used.

We have also argued that we can consider the tolerance/invariance of a neuron in terms of endo-origin curvature. We argue that the increased selectivity and tolerance of higher level visual neurons (Rust & DiCarlo (2010)) implies a combination of high exo-origin curvature and high endo-origin curvature. We believe we can use this approach to analyze current networks available in the literature (e.g., Karklin & Lewicki (2005); Serre et al. (2007); Yamins et al. (2014)). We are also currently using the ideas in this paper to develop networks that learn to be both selective and tolerant based on the statistical structure of natural scenes.

8. Acknowledgment

This material is based upon work supported by Google Faculty Research Award to David J. Field. Some of the early work was generously supported by NEI grant EY12241 to Professor Jack Gallant while Michael Wu and David Field were at the University of California Berkeley in 2003. Michael Wu was also supported by the CSGF program funded by the US Department of Energy. We thank Bruno Olshausen for helpful comments on an earlier version of manuscript, and Christoff Zetzsche for his comments as a reviewer.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *JOSA A*, *2*, 284–299.

Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, *7*, 531–546.

Albrecht, D. G., Geisler, W. S., & Crane, A. M. (2003). Nonlinear properties of visual cortex neurons: Temporal dynamics, stimulus selectivity, neural performance. In L. Chalupa, & J. Werner (Eds.), *The Visual Neurosciences* The Visual Neurosciences (pp. 825–837). Boston: MIT Press.

Albrecht, D. G., Geisler, W. S., Frazor, R. A., & Crane, A. M. (2002). Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *Journal of Neurophysiology*, *88*, 888–913.

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *J Neurophysiol*, *48*, 217–237.

Bell, A. J., & Sejnowski, T. J. (1997). The independent components of natural scenes are edge filters. *Vision research*, *37*, 3327–3338.

Berkes, P., & Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision*, *5*, 9.

Bishop, P., Coombs, J. S., & Henry, G. (1973). Receptive fields of simple cells in the cat striate cortex. *The Journal of physiology*, *231*, 31–60.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *J. of Physiol.*, *203*, 237–260.

Brady, N., & Field, D. J. (2000). Local contrast in natural images: normalisation and coding efficiency. *PERCEPTION-LONDON-*, *29*, 1041–1056.

Chandler, D. M., & Field, D. J. (2007). Estimates of the information content and dimensionality of natural scenes from proximity distributions. *JOSA A*, *24*, 922–941.

Dean, A., & Tolhurst, D. (1983). On the distinctness of simple and complex cells in the visual cortex of the cat. *The Journal of Physiology*, *344*, 305–325.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences*, *11*, 333–341.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*, 415–434.

EinhaÈuser, W., Kayser, C., KoÈnig, P., & KoÈrding, K. P. (2002). Learning the invariance properties of complex cells from their responses to natural stimuli. *European Journal of Neuroscience*, *15*, 475–486.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, *4*, 2379–2394.

Field, D. J. (1994). What is the goal of sensory coding? *Neural computation*, *6*, 559–601.

Field, D. J., & Wu, M. (2004). An attempt towards a unified account of non-linearities in visual neurons. *Journal of vision*, *4*, 283–283.

Golden, J. R. (2015). *A unified approch to the non-linearities of visual neurons: the curved geometry of neural response surfaces.*. Ph.D. thesis Cornell University.

Golden, J. R., Vilankar, K. P., & Field, D. J. (2016). Measuring selectivity and tolerance in networks using principal curvatures.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181–197.

Horwitz, G. D., & Hass, C. A. (2012). Nonlinear analysis of macaque v1 color tuning reveals cardinal directions for cortical color processing. *Nature neuroscience*, *15*, 913–919.

Hosseini, R., Sinz, F., & Bethge, M. (2010). Lower bounds on the redundancy of natural images. *Vision research*, *50*, 2213–2222.

Hoyer, P. O. (2003). Modeling receptive fields with non-negative sparse coding. *Neurocomputing*, *52*, 547–552.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J. Physiol.*, *160*, 106–154.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on*, *10*, 626–634.

Hyvärinen, A., & Hoyer, P. O. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vis. Res.*, *41*, 2413–2423.

Kagan, I., Gur, M., & Snodderly, D. M. (2002). Spatial organization of receptive fields of v1 neurons of alert monkeys: comparison with responses to gratings. *Journal of neurophysiology*, *88*, 2557–2574.

Karklin, Y., & Lewicki, M. S. (2005). A hierarchical bayesian model for learning nonlinear statistical regularities in nonstationary natural signals. *Neural Computation*, *17*, 397–423.

Körding, K. P., Kayser, C., & König, P. (2003). On the choice of a sparse prior. *Reviews in the Neurosciences*, *14*, 53–62.

Laughlin, S. B. (1981). A simple coding procedure enhances a neurons information capacity. *Z. Naturforsch*, *36*, 51.

Le, Q. V., Karpenko, A., Ngiam, J., & Ng, A. Y. (2011a). Ica with reconstruction cost for efficient overcomplete feature learning. In *Advances in Neural Information Processing Systems* (pp. 1017–1025).

Le, Q. V., Zou, W. Y., Yeung, S. Y., & Ng, A. Y. (2011b). Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (pp. 3361–3368). IEEE.

Lee, H., Battle, A., Raina, R., & Ng, A. Y. (2006). Efficient sparse coding algorithms. In *Advances in neural information processing systems* (pp. 801–808).

Miao, X., & Rao, R. P. (2007). Learning the lie groups of visual invariance. *Neural computation*, *19*, 2665–2693.

Morrone, M. C., Burr, D., & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. i. neurophysiological evidence. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *216*, 335–354.

Ohzawa, I., Sclar, G., & Freeman, R. D. (1985). Contrast gain control in the cats visual system. *J Neurophysiol*, *54*, 651–667.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vis. Res.*, *37*, 3311–3325.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding v1? *Neural computation*, *17*, 1665–1699.

Prenger, R., Wu, M. C.-K., David, S. V., & Gallant, J. L. (2004). Nonlinear v1 responses to natural scenes revealed by neural network analysis. *Neural Networks*, *17*, 663–679.

Rao, R. P., & Ruderman, D. L. (1999). Learning lie groups for invariant visual perception. *Advances in neural information processing systems*, (pp. 810–816).

Rifai, S., Mesnil, G., Vincent, P., Muller, X., Bengio, Y., Dauphin, Y., & Glorot, X. (2011). Higher order contractive auto-encoder. In *Machine Learning and Knowledge Discovery in Databases* (pp. 645–660). Springer.

Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance (invariance) both increase as visual information propagates from cortical area v4 to it. *The Journal of Neuroscience*, *30*, 12978–12995.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, *4*, 819–825.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *29*, 411–426.

Skottun, B. C., De Valois, R. L., Grosof, D. H., Movshon, J. A., Albrecht, D. G., & Bonds, A. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision research*, *31*, 1078–1086.

Tadmor, Y., & Tolhurst, D. (2000). Calculating the contrasts that retinal ganglion cells and lgn neurones encounter in natural scenes. *Vision research*, *40*, 3145–3157.

Tolhurst, D., & Heeger, D. (1997). Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. *Visual neuroscience*, *14*, 293–309.

Touryan, J., Felsen, G., & Dan, Y. (2005). Spatial structure of complex cell receptive fields measured with natural images. *Neuron*, *45*, 781–791.

Tsai, C.-Y., & Cox, D. D. (2015). Measuring and understanding sensory representations within deep networks using a numerical optimization framework. *arXiv preprint arXiv:1502.04972*, .

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*, 1273–1276.

Wang, J., Sohl-Dickstein, J., & Olshausen, B. (2009). Unsupervised learning of lie group operators from image sequences. In *Frontiers in Systems Neuroscience. Conference Abstract: Computational and systems neuroscience*.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, (p. 201403112).

Yu, C., & Levi, D. M. (1997). Spatial facilitation predicted with end-stopped spatial filters. *Vision Research*, *37*, 3117–3127.

Zetzsche, C., Krieger, G., & Wegmann, B. (1999). The atoms of vision: Cartesian or polar? *JOSA A*, *16*, 1554–1565.

Zetzsche, C., & Nuding, U. (2005). Nonlinear and higher-order approaches to the encoding of natural scenes. *Network: Computation in Neural Systems*, *16*, 191–221.

Zetzsche, C., & Rohrbein, F. (2001). Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network: Computation in Neural Systems*, *12*, 331–350.

Zhu, M., & Rozell, C. J. (2013). Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. *PLoS computational biology*, *9*, e1003191.