HAL
open science

A Differential Model of the Complex Cell

Miles Hansard, Radu Horaud

To cite this version:

Miles Hansard, Radu Horaud. A Differential Model of the Complex Cell. Neural Computation, 2011, 23 (9), pp.2324-2357. 10.1162/NECO_a_00163 . inria-00590266

HAL Id: inria-00590266

https://inria.hal.science/inria-00590266

Submitted on 10 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LETTER Communicated by Friedrich Sommer

A Differential Model of the Complex Cell

Miles Hansard
miles.hansard@inrialpes.fr
Radu Horaud
radu.horaud@inrialpes.fr
Inria Rhône-Alpes, Montbonnot, France 38330

The receptive fields of simple cells in the visual cortex can be understood as linear filters. These filters can be modeled by Gabor functions or gaussian derivatives. Gabor functions can also be combined in an energy model of the complex cell response. This letter proposes an alternative model of the complex cell, based on gaussian derivatives. It is most important to account for the insensitivity of the complex response to small shifts of the image. The new model uses a linear combination of the first few derivative filters, at a single position, to approximate the first derivative filter, at a series of adjacent positions. The maximum response, over all positions, gives a signal that is insensitive to small shifts of the image. This model, unlike previous approaches, is based on the scale space theory of visual processing. In particular, the complex cell is built from filters that respond to the 2D differential structure of the image. The computational aspects of the new model are studied in one and two dimensions, using the steerability of the gaussian derivatives. The response of the model to basic images, such as edges and gratings, is derived formally. The response to natural images is also evaluated, using statistical measures of shift insensitivity. The neural implementation and predictions of the model are discussed.

1 Introduction

It is useful to distinguish between simple and complex cells in the visual cortex, as Hubel and Wiesel (1962) proposed. The original classification is based on four characteristic properties of simple cells, as follows. First, the receptive field has distinct excitatory and inhibitory subregions. Second, there is spatial summation within each subregion. Third, there is antagonism between the excitatory and inhibitory subregions. Fourth, the response to any stimulus can be predicted from the receptive field map. Cells that do not show these characteristics can be classified, by convention, as complex.

The distinction between simple and complex cells has endured, subject to certain qualifications (although an alternative view is described by Mechler & Ringach, 2002). In particular, it has been argued that the principal

 © 2011 Massachusetts Institute of Technology

characteristic of complex cells is the phase invariance of the response (Carandini, 2006; Movshon, Thompson, & Tolhurst, 1978a). This means that a complex cell, which is tuned to a particular orientation and spatial frequency, is not sensitive to the precise location of the stimulus within the receptive field (Kjaer, Gawne, Hertz, & Richmond, 1997; Mechler, Reich, & Victor, 2002). If the stimulus consists of a moving (or flickering) grating, then phase invariance can be quantified by the relative modulation a_1/a_0, where a_1 is the amplitude associated with the fundamental frequency of the response and a_0 is the mean response (De Valois, Albrecht, & Thorell, 1982; Movshon et al., 1978a; Skottun et al., 1991). If this ratio is less than one, the cell can be classified as complex.

The standard model of the simple cell is based on a linear filter that is localized in position, spatial frequency, and orientation. The output of this filter is subject to a nonlinearity, such as squaring, and may also be normalized by the responses of nearby simple cells. The theoretical framework of this model is well advanced (as reviewed by Carandini et al., 2005; Dayan & Abbott, 2001), owing to the assumed linearity of the underlying spatial filters. Although the physiology of the complex cell is increasingly well understood (as reviewed by Alonso & Martinez, 1998; Martinez & Alonso, 2003; Spitzer & Hochstein, 1988), the appropriate theoretical framework is less clear. It is useful to adopt the distinction between position and phase models that was made by Fleet, Wagner, and Heeger (1996) in the analysis of binocular processing. It is invariance to position or phase that is of interest in the context here.

A position-invariant model of the complex cell can be constructed from a set of simple cells, as described by Hubel and Wiesel (1962). Suppose, for example, that the simple cells are represented by linear filters of a common orientation but different spatial positions. If the responses of these filters are summed, then the corresponding complex cell will be tuned to an oriented element that appears anywhere in the union of the simple cell receptive fields (Spitzer & Hochstein, 1988). This scheme is the basis of the subunit model, which was further developed and tested by (Movshon et al., 1978a; Movshon, Thompson, & Tolhurst, 1978b). The results of these experiments are consistent with the idea that the complex response is based on a group of spatially linear subunits. The most straightforward way to combine the individual responses would be by rectifying and summing them, but the maximum could also be taken (Riesenhuber & Poggio, 1999). The main disadvantage of the subunit model is that it is too general in its basic form. For example, it allows arbitrarily complicated receptive fields to be constructed, as there is no intrinsic constraint on the positions, orientations, or spatial frequencies of the subunits.

Alternatively, a phase-invariant model can be constructed from a pair of filters of different shapes (Adelson & Bergen, 1985). If the filters have a quadrature relationship, then their responses to a sinusoidal stimulus will differ in phase by $\pi/2$. It follows that the sum of the squared outputs will be

invariant to the phase of the input. The application of this energy model to spatial vision (Atherton, 2002; Emerson, Bergen, & Adelson, 1992; Morrone & Burr, 1988; Wundrich, von der Malsburg, & Würtz, 2004) is motivated by the observed phase differences in the receptive fields of adjacent simple cells (Pollen & Ronner, 1981). Indeed, these RFs can be well represented by odd and even Gabor functions (Daugman, 1985; Jones & Palmer, 1987; Pollen & Ronner, 1983). There are two problems with the energy model in the present context. First, there is no generally agreed way to combine energy mechanisms across different frequencies and orientations (one approach is described by Fleet et al., 1996). This is an obstacle to the construction of mechanisms that show more complicated invariances, such as those found in areas MT and MST (Orban, 2008). Second, the quadrature filters that are best suited to the energy model are not convenient for the general description of 2D image-structure. The concept of phase itself becomes somewhat complicated in two or more dimensions (Felsberg & Sommer, 2001), and the quadrature representation of more complex image-features, such as edge curvature, is indirect. It must be emphasized that 2D images contain important structures (e.g., luminance saddle points) that have no analog in the 1D signals to which the energy model is ideally suited. A realistic framework for spatial vision must be capable of representing the full variety of 2D structures (Ben-Shahar & Zucker, 2004; Dobbins, Zucker, & Cynader, 1987; Petitot, 2003).

This work is motivated by the difficulty of extending the energy model to more complicated 2D image-features and spatial transformations. These extensions require a representation of the local image geometry rather than the local phase and frequency structure. The new approach, like the energy model, is based on a set of odd and even linear filters at a common location. The outputs of these filters are nonlinearly combined, again as in the energy model. The combination, however, involves the implicit construction of spatially offset subunits. The differential model, in this sense, can be seen as a reformulation of the Hubel and Wiesel (1962) scheme.

1.1 Formal Overview. A minimal overview of the new model will now be given. Let $S(\boldsymbol{x})$ be the original scalar image, where $\boldsymbol{x} = (x, y)^\top$, and consider a spatial array of simple cells, parameterized by preferred frequency and orientation. These receptive fields will be modeled by kth order directional derivatives of the gaussian kernel, $G_k(\boldsymbol{x}, \sigma, \theta) = (\boldsymbol{v} \cdot \nabla)^k G_0(\boldsymbol{x}, \sigma)$, where σ is the spatial scale, θ is the orientation, and $\boldsymbol{v} = (\cos\theta, \sin\theta)^\top$. The simple cell representation $S_k(\boldsymbol{x}, \sigma, \theta)$ is given by the convolution of these filters with the image

$$S_k(\boldsymbol{x}, \sigma, \theta) = G_k(\boldsymbol{x}, \sigma, \theta) \star S(\boldsymbol{x}). \tag{1.1}$$

In particular, consider the magnitude of the first derivative signal, $|S_1(\boldsymbol{x}, \sigma, \theta)|$. This will be large if there is a step-like edge at $\boldsymbol{x}$, with the

luminance boundary perpendicular to $\boldsymbol{v}$. Now suppose that the edge is shifted by some amount in direction $\boldsymbol{v}$. This means that the magnitude $|S_1(\boldsymbol{x}, \sigma, \theta)|$ will fall, but the nonlinear function,

$$C(\boldsymbol{x}, \sigma, \theta) = \max_t \left|S_1(\boldsymbol{x} + t\boldsymbol{v}, \sigma, \theta)\right|, \quad \text{where} \quad |t| \leq \rho, \tag{1.2}$$

will remain large unless the shift exceeds the range ρ. Equations 1.1 and 1.2 will be the basic models of simple cells $S_k(\boldsymbol{x}, \sigma, \theta)$ and complex cells $C(\boldsymbol{x}, \sigma, \theta)$ in this letter (full derivations are given in section 2). The complex cell, which inherits the scale and orientation tuning (σ, θ), has a receptive field of radius ρ, centered on position $\boldsymbol{x}$. It can be seen that equation 1.2 is just a special case of the Hubel and Wiesel (1962) subunit model, with simple cells distributed along the spatial axis $\boldsymbol{v}$, and "max" being the combination rule. It has already been argued that this model is too general. For example, there is no natural limit on the size ρ of the complex receptive field in equation 1.2.

Suppose, however, that access to the first-order directional structure around position $\boldsymbol{x}$ is replaced by access to the higher-order directional structure *at* position $\boldsymbol{x}$. Mathematically, this means that the function $S_1(\boldsymbol{x} + t\boldsymbol{v}, \sigma, \theta)$ of the scalar t is replaced by the values $S_k(\boldsymbol{x}, \sigma, \theta)$ indexed by $k = 1, \ldots, K$. This is interesting for three reasons. First, the model becomes inherently local, because the filters G_k that compute the values S_k are now centered at the same point $\boldsymbol{x}$. Second, the filters G_k are symmetric or antisymmetric about the point $\boldsymbol{x}$ and resemble the Gabor functions used in the energy model. Third, the values S_k can be obtained from a linear transformation of the Kth order local jet at $\boldsymbol{x}$, so this scheme is compatible with the scale-space theory described above.

To be more specific, we will show that the first-order structure in the neighborhood $\boldsymbol{x} + t\boldsymbol{v}$, as in equation 1.2, can be estimated from a linear combination of the directional derivatives S_k,

$$S_1(\boldsymbol{x} + t\boldsymbol{v}, \sigma, \theta) \approx \sum_{k=1}^{K} P_k(t) S_k(\boldsymbol{x}, \sigma, \theta), \tag{1.3}$$

where the functions $P_k(t)$ are fixed polynomials. This approximation will then be substituted into the right-hand side of equation 1.2. It will be shown in section 2.2 that the approximation, equation 1.3, can be motivated by a Maclaurin expansion in powers of t. This can also be interpreted, as shown in Figure 1, as the synthesis of spatially offset filters, using the gaussian derivatives as a basis. A matrix formulation of this model will be given in section 2.3. An optimal (and image-independent) construction of the polynomials $P_k(t)$ will be given in section 2.4. The case in which the derivatives on the right-hand side of equation 1.3 are in another direction, $\phi \neq \theta$, is treated in section 2.5.

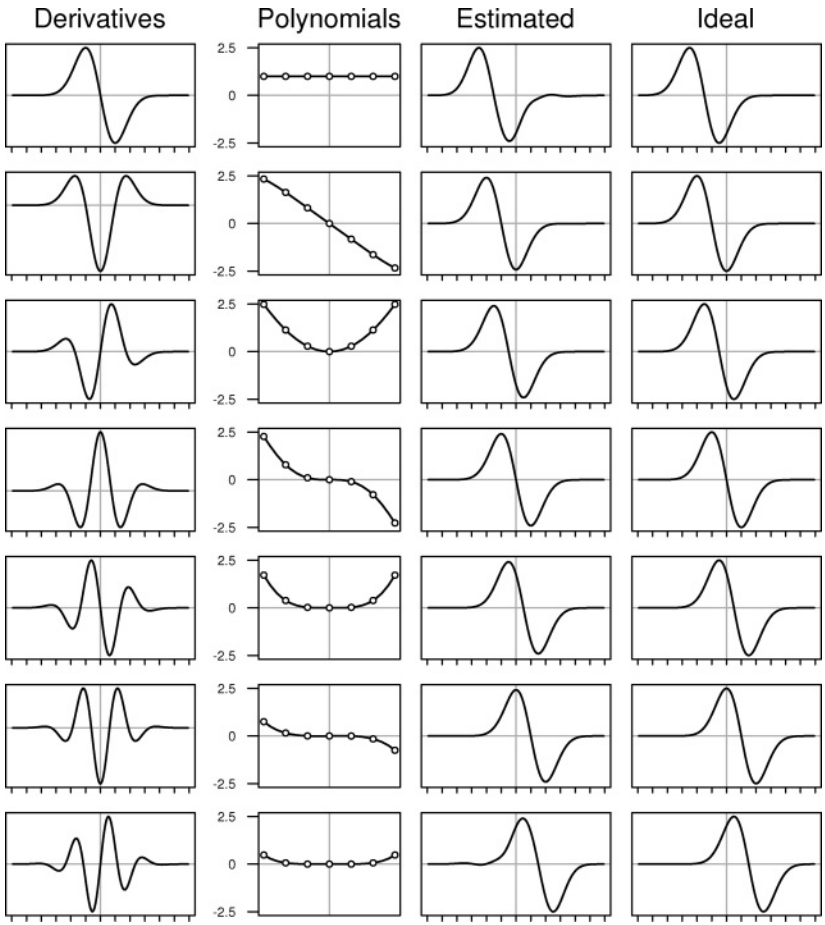

Figure 1: Construction of offset filters. (Column 1) The gaussian derivatives $G_k(x, \sigma)$, scaled for display, of orders $1, \ldots, K$, where $K = 7$. (Column 2) The corresponding polynomial interpolation functions $P_k(t)$, of order $K - 1$. Note that $P_k(t)$ resembles the monomial $t^{(k-1)}$. (Column 3) Estimated filters, $F(t_j, x)$, which are offset versions of $G_1(x, \sigma)$. (Column 4) Ideal filters $F_\star(t_j, x)$. The synthesis equation is $F(t_j, x) = \sum_k P_k(t_j) G_k(x, \sigma)$, where each weight $P_k(t_j)$ corresponds to the jth dot on the kth polynomial in column 2.

1.2 Contributions and Organization. The model presented in this letter is quite different from the previous approaches. The main contribution is a "differential" model of the complex cell, which is exactly steerable and fits naturally into the geometric approach to image analysis (Koenderink & van Doorn, 1987, 1990). This shows that it is possible to analyze the local

image geometry and obtain a shift-invariant response using a common set of filters.

The new model is developed in section 2, using linear algebra and least-squares optimization. The accuracy of the estimated filters is evaluated in section 3. This section also derives the exact response of the ideal filters to several basic stimuli. Some preliminary experiments on natural images are reported. The biological interpretation of the model, and its predictions, are discussed in section 4. Future directions, and the relationship of the model to scale-space theory, are also discussed.

2 Differential Model

The following notation will be used here. Matrices and vectors are written in bold, for example, $\boldsymbol{M}$, $\boldsymbol{v}$, where $\boldsymbol{M}^\top$ is the transpose and $\boldsymbol{M}^+$ is the Moore-Penrose inverse (Press, Teukolsky, Vetterling, & Flannery, 1992). The convolution of functions is $F(x) \star G(x) = \int_{-\infty}^{\infty} F(x-y)\, G(y)\, \mathrm{d}y$. Some properties of the gaussian derivatives $G_k(x, \sigma)$ will now be reviewed. There is no particular spatial scale at which a natural image should be analyzed. It is therefore desirable to represent the image in a scale space, so that a range of resolutions can be considered (Koenderink, 1984). The preferred way to do this is by convolution with a gaussian kernel. It follows that the structure of the image, at a given scale, can be analyzed by the spatial derivatives of the corresponding gaussian. The kth order derivatives of a 1D gaussian, $G_k = \mathrm{d}^k/\mathrm{d}x^k\, G_0$ can be expressed as

$$G_k(x, \sigma) = \left(\frac{-1}{\sigma\sqrt{2}}\right)^k H_k\left(\frac{x}{\sigma\sqrt{2}}\right) G_0(x, \sigma), \tag{2.1}$$

$$G_0(x, \sigma) = \exp\left(\frac{-x^2}{2\sigma^2}\right), \tag{2.2}$$

where $G_0(x, \sigma)$ is the original gaussian, k is a positive integer, and $H_k(x)$ is the kth Hermite polynomial. The first seven gaussian derivatives are shown in column 1 of Figure 1. It will also be useful to introduce two normalizations of the gaussian derivatives,

$$G_k^0(x, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, G_k(x, \sigma) \quad \text{and} \quad G_k^1(x, \sigma) = \frac{1}{2}\, G_k(x, \sigma), \tag{2.3}$$

which are defined so that $\int |G_k^k(x)| \mathrm{d}x = 1$. In particular, G_0^0 and G_1^1 are the L^1-normalized blurring and differentiating filters, respectively. This superscript notation will not be used unless a particular normalization is important (e.g., in section 3.2).

The two-dimensional gaussian derivative, in direction θ with $\boldsymbol{x} = (x, y)^\top$, will be written $G_k(\boldsymbol{x}, \sigma, \theta) = (\boldsymbol{v} \cdot \nabla)^k G_0(\boldsymbol{x}, \sigma)$, as in section 1.1. Two special properties of these filters should be noted. First, the filter $G_k(\boldsymbol{x}, \sigma, \theta)$ is separable in the local coordinate system that is defined by the direction of differentiation. This means that the 2D filter can be obtained from the product of 1D filters $G_k(x_\theta, \sigma)$ and $G_0(y_\theta, \sigma)$. Second, the gaussian derivatives are steerable, meaning that $G_k(\boldsymbol{x}, \sigma, \theta)$ can be obtained from a linear combination of derivatives in other directions, $G_k(\boldsymbol{x}, \sigma, \phi_j)$, where $j = 1, \ldots, k+1$. These facts make it possible to analyze a multidimensional filter, in many cases, in terms of 1D functions (Freeman & Adelson, 1991).

The first derivative, G_1, will be used as the basic model of a complex subunit (which is also a simple cell-receptive field). This choice is motivated by two observations. First, it is well established that gradient filters can be used to detect edges, as well as more complex image features (Canny, 1986; Harris & Stephens, 1988). Second, G_1 is the first zero-mean filter in the local-jet representation of the image, which is physiologically and mathematically convenient. The extension to higher-order subunits is straightforward, as discussed in section 4.3.

2.1 Filter Arrays. This section puts the system of simple cells, introduced in section 1.1, into a standard signal processing framework. This will be done in 1D in order to simplify the notation. The extension to 2D is straightforward. The 1D version of the simple cell response, equation 1.1 is $S_k(u, \sigma) = \int_{-\infty}^{\infty} G_k(u - x, \sigma) S(x) \mathrm{d}x$. If $k = 1$ and the filter G_1 is offset by an amount t, then the convolution can be expressed as

$$\begin{aligned} S_1(u - t, \sigma) &= \int_{-\infty}^{\infty} G_1(u - t - x, \sigma)\, S(x)\, \mathrm{d}x \\ &= -\int_{-\infty}^{\infty} G_1(x - t, \sigma)\, S(x - u)\, \mathrm{d}x. \end{aligned} \tag{2.4}$$

Here the antisymmetry of $G_1(x, \sigma)$ has been used to show that the result is equal to the negative correlation of the filter and signal. It is evident that if t could be kept equal to u, then the signal shift would have no effect on the result. The prototypical stimulus for G_1 is the step-edge $S(x) = \mathrm{sgn}(x)$. In this case, the filter and signal are anticorrelated, and so the final response, equation 2.4, is nonnegative.

2.2 Functional Representation. It was established in section 2.1 that the response $S_1(u - t, \sigma)$ can be constructed from the offset filters $G_1(x - t, \sigma)$. This means that the desired approximation, equation 1.3, can be treated as a filter design problem. The following notation will be adopted for the offset

filters:

$$\begin{aligned} F(0,x) &= G_1(x,\sigma), \\ F(t,x) &\approx G_1(x-t,\sigma), \end{aligned} \tag{2.5}$$

which also depend on the spatial scale and derivative order, but it will not be necessary to make this explicit in the notation. It will suffice to analyze a single filter that, without loss of generality, is located at the origin $\boldsymbol{x} = (0,0)^\top$ of the spatial coordinate system. The linear response of this filter is defined in relation to equation 2.4 as

$$R(t,u) = -\int_{-\infty}^{\infty} F(t,x)\, S(x-u)\, \mathrm{d}x. \tag{2.6}$$

The complex response at $\boldsymbol{x} = (0,0)^\top$, with reference to equation 1.2, can now be expressed as a function of the signal translation u;

$$C(u) = \max_t \left| R(t,u) \right| \quad \text{where} \quad |t| \le \rho. \tag{2.7}$$

The actual value of u in general has no particular significance. It will be more important to consider the response $R(t,u)$ as u changes. In particular, suppose that $|R(t,u)|$ is high at the stimulus position $u = u_0$. If the response is insensitive to a slight translation of the signal, then $\partial^2 C / \partial u^2 \approx 0$ at u_0.

The approximation problem in equation 2.5 will now be addressed. The filter $F(t,x)$ can be defined in relation to the Maclaurin expansion of $G_1(x - t,\sigma)$ with respect to the offset t. If image derivatives up to order K are available, then the approximation is

$$F(t,x) = \sum_{k=0}^{K-1} \frac{(-t)^k}{k!}\, G_{k+1}(x,\sigma). \tag{2.8}$$

The key observation is that the filters $G_{k+1}(x,\sigma)$ in equation 2.8 are precisely those that compute the local jet coefficients, of order $1,\ldots,K$, at the point $x = 0$. In other words, the family of shifted filters $F(t,x)$ has been obtained from the family of non shifted derivatives $G_k(x,\sigma)$. It can be seen from equation 2.1, 2.2, and 2.8 that the estimated function $F(t,x)$ decreases to 0 for large $|x|$, owing to the exponential tails of G_0.

The definition (see equation 2.8) is usable in practice, as will be shown in section 3.1. There are, however, two difficulties with the scheme described. First, although $F(t,x)$ is an approximation of $G_1(x-t,\sigma)$, the nature of this approximation (a polynomial with the given derivatives) may be inappropriate. Second, as expected, the approximation 2.8 is not well behaved for

large $|t|$. Both of these problems can be addressed by replacing the Maclaurin series with a more flexible construction of $F(t, x)$. This is done by substituting a general polynomial $P_k(t)$ in place of each monomial $(-t)^k/k!$ in equation 2.8, leading to

$$F(t, x) = \sum_{k=1}^{K} P_k(t)\, G_k(x, \sigma). \tag{2.9}$$

The K polynomials $P_k(t)$ are constructed from standard monomial basis functions t^j and coefficients c_{jk}. The order of each polynomial will be $K - 1$, for consistency with the original series approximation, equation 2.8. It follows that the polynomials are

$$P_k(t) = \sum_{j=0}^{K-1} t^j c_{jk} \quad \text{where} \quad 1 < k \leq K. \tag{2.10}$$

The problem has now been altered to that of finding K^2 appropriate coefficients c_{jk}. This will be treated, in sections 2.4 and 2.5, as the optimization of

$$\arg\min_{\boldsymbol{C}} \iint \left|F(t, x) - G_1(x - t, \sigma)\right|^2 \mathrm{d}x\, \mathrm{d}t,$$

where $F(t, x)$ is the family of filters defined in equation 2.9 and $\boldsymbol{C}$ is the matrix of coefficients c_{jk}. This optimization scheme generalizes immediately to filters in any number of dimensions. The simple Maclaurin scheme, equation 2.8, remains a useful model, because the optimal polynomials are, in practice, close to the original monomials $P_k(t) \approx t^{(k-1)}$, as can be seen in Figure 1.

It is important to note that once the coefficients c_{jk} have been estimated, the location of the synthetic filter $F(t, x)$ can be varied continuously with respect to the offset t. Any set of translated filters $F(t_i, x)$ can be obtained, provided $|t_i| \leq \rho$ for $i = 1, \ldots M$, by resampling the monomial basis functions as t_i^j and, then repeating equations 2.9 and 2.10. Furthermore, the principle of shiftability states that the convolution $f(x - t) \star s(x)$ can be represented in a finite basis $f(x - t_i)$, provided that the filter f is band limited (Perona, 1995; Simoncelli, Freeman, Adelson, & Heeger, 1992). The filters $F(t, x)$ are not band limited, but they do decay exponentially, as the Fourier transforms have a gaussian factor. This means, in practice, that the linear response $R(t, u)$ in equation 2.6 can be represented by a suitable discretization $R(t_i, u)$, where the shift resolution $\Delta t = 2\rho/(M - 1)$ can be chosen to achieve any desired accuracy.

2.3 Matrix Representation. It will be convenient to represent the filter construction in terms of matrices. This results in a compact formulation and prepares for the least-squares estimation procedure that will be introduced in section 2.4. Suppose that M filters, each of length N, are to be constructed and that the highest available derivative is of order K. Each filter will be represented as a row vector, so that the collection of offset filters forms an $M \times N$ matrix $\boldsymbol{F}$. Note that this representation applies in any number of dimensions, provided that the positions of the filter samples are consistently identified with the column indices of $\boldsymbol{F}$.

The columns of another matrix $\boldsymbol{P}$ will contain the K polynomials $P_k(t)$ from equation 2.10. These polynomials must be sampled at M points t_i; hence, $\boldsymbol{P}$ has dimensions $M \times K$. Let the sampled monomial basis functions t_i^j be the columns of the matrix $\boldsymbol{B}$, which therefore must have the same dimensions, $M \times K$. The monomials are weighted by the $K \times K$ matrix of coefficients $\boldsymbol{C}$, such that

$$\boldsymbol{P} = \boldsymbol{B}\boldsymbol{C}, \tag{2.11}$$

where column k of $\boldsymbol{C}$ contains the coefficients of P_k. Let each row of the $K \times N$ matrix $\boldsymbol{G}$ contain the sampled gaussian derivative $G_k(x_i, \sigma)$. Each offset filter should be a linear combination of the gaussian derivatives, constructed from the polynomials P_k. It follows from equation 2.9 that

$$\boldsymbol{F} = \boldsymbol{P}\boldsymbol{G}. \tag{2.12}$$

Let the column vector $\boldsymbol{s}$ contain the sampled signal, $\boldsymbol{s} = \big(S(x_1), \ldots, S(x_N)\big)^{\top}$. This means that the response vector $\boldsymbol{r} = \big(R(t_1, u), \ldots, R(t_M, u)\big)^{\top}$ is obtained according to equations 2.6 and 2.12 as

$$\begin{aligned} \boldsymbol{r} &= -\boldsymbol{F}\boldsymbol{s} \\ &= -\boldsymbol{P}\boldsymbol{G}\boldsymbol{s}. \end{aligned} \tag{2.13}$$

This clearly shows that the response $\boldsymbol{r}$ is simply a linear transformation $\boldsymbol{P}$ of the Kth order gaussian jet, $\boldsymbol{G}\boldsymbol{s} = \big(S_1(x, \sigma), \ldots, S_K(x, \sigma)\big)^{\top}$. The implication is that the filter bank $\boldsymbol{F}$ need not be explicitly constructed; rather, the response $\boldsymbol{r}$ is computed directly from the K image derivatives S_k at $\boldsymbol{x}$.

2.4 Unconstrained Estimation. It will now be shown that the K^2 unknown coefficients, contained in the matrix $\boldsymbol{C}$, can be obtained by standard least-squares methods. It should be emphasized that this is a filter design problem; the matrix $\boldsymbol{P}$ is fixed for all signals, and the response $\boldsymbol{r}$ is obtained according to equation 2.13.

Let the $M \times N$ matrix $\boldsymbol{F}_\star$ contain the true derivative filters, such that the ijth element is $G_1(x_j - t_i, \sigma)$. The approximation $\boldsymbol{F} \approx \boldsymbol{F}_\star$ can be expressed, according to equations 2.11 and 2.12, as the product

$$\boldsymbol{F}_\star \approx \boldsymbol{B}\boldsymbol{C}\boldsymbol{G}, \quad \text{of which} \quad \boldsymbol{C} = \boldsymbol{B}^+ \boldsymbol{F}_\star \boldsymbol{G}^+ \tag{2.14}$$

is the solution in the least-squares sense. This formulation requires two Moore-Penrose inverses, which can be computed from the singular-value decompositions of the monomial basis and gaussian derivative matrices $\boldsymbol{B}$ and $\boldsymbol{G}$, respectively. It is, however, more efficient to solve this problem using QR decompositions, as follows. There are, in practice, more offsets than derivative filters $M > K$, as well as more spatial samples than derivative filters $N > K$. The basis matrix has full column rank K and can be factored as $\boldsymbol{B} = \boldsymbol{Q}_B \boldsymbol{R}_B$. The derivative matrix has full row rank K, and so its transpose can be factored as $\boldsymbol{G}^\top = \boldsymbol{Q}_G \boldsymbol{R}_G$. It follows that the solution 2.14 can be obtained via $\boldsymbol{B}^+ = \boldsymbol{R}_B^{-1} \boldsymbol{Q}_B^\top$ and $\boldsymbol{G}^+ = \boldsymbol{Q}_G \boldsymbol{R}_G^{-\top}$.

2.5 Constrained Estimation. The least-squares construction of the filters $\boldsymbol{F}$ was described in the preceding section. The method is quite usable but has two shortcomings. First, if one of the shifts t_i is 0, then $F(0, x) \approx G_1(x, \sigma)$, but it would be preferable to make this an exact equality, so that the original filter is returned as in equation 2.5. The second shortcoming of the method in section 2.4 is that, in two or more dimensions, the orientation of the derivative filters in the basis set $\boldsymbol{G}$ may not match that of the target set $\boldsymbol{F}_\star$. Both of these problems will be solved below.

2.5.1 Additive Solution. The requirement $F(0, x) = G_1(x, \sigma)$ is satisfied by an additive model in which the polynomial $P_1(t)$ that weights $G_1(x, \sigma)$ is always unity, and all other polynomials pass through 0 when $t = 0$. This implies the following partitioning of the derivative, monomial, and coefficient matrices,

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{g}_1^\top \\ \boldsymbol{G}_\Delta \end{pmatrix} \quad \boldsymbol{B} = (\boldsymbol{1} \quad \boldsymbol{B}_\Delta) \quad \boldsymbol{C} = \begin{pmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{C}_\Delta \end{pmatrix}, \tag{2.15}$$

where $\boldsymbol{1}$ is the column vector of M 1's and $\boldsymbol{0}$ is the column vector of $(K-1)$ 0's. The $1 \times N$ vector $\boldsymbol{g}_1^\top$ contains the first derivative filter $G_1(x, \sigma)$, while the $(K-1) \times N$ matrix $\boldsymbol{G}_\Delta$ contains the higher-order filters. The columns of the $M \times (K-1)$ matrix $\boldsymbol{B}_\Delta$ contain the sampled monomials, excluding the constant vector $\boldsymbol{1}$. The unknown matrix $\boldsymbol{C}$ will be recovered in the form indicated, where $\boldsymbol{C}_\Delta$ has dimensions $(K-1) \times (K-1)$.

The product of $\boldsymbol{B}$ and $\boldsymbol{C}$, as in equation 2.11, now gives $\boldsymbol{P} = (\boldsymbol{1} \ \ \boldsymbol{P}_\Delta)$, where the columns of $\boldsymbol{P}_\Delta = \boldsymbol{B}_\Delta \boldsymbol{C}_\Delta$ are polynomials without constant terms.

It follows that the product $\boldsymbol{PG}$, as in equation 2.12, gives the additive approximation,

$$\boldsymbol{F}_\star \approx \boldsymbol{1}\boldsymbol{g}_1^\top + \boldsymbol{B}_\Delta \boldsymbol{C}_\Delta \boldsymbol{G}_\Delta, \tag{2.16}$$

where $\boldsymbol{1}\boldsymbol{g}_1^\top$ is the rank-one matrix containing M identical rows $\boldsymbol{g}_1^\top$. Note that if the ith row of $\boldsymbol{F}_\star$ corresponds to $t = 0$, then the ith row of $\boldsymbol{B}_\Delta \boldsymbol{C}_\Delta$ must be 0, this being the evaluation of the polynomials $\sum_{j=1}^{K-1} t^j c_{jk}$ at $t = 0$. It follows that the ith row of $\boldsymbol{F}_\star$ is exactly recovered from equation 2.16 as $\boldsymbol{g}_1^\top$, and so the constraint $F(0, x) = G_1(x, \sigma)$ has been imposed. The unknown coefficients $\boldsymbol{C}_\Delta$ are recovered by subtracting $\boldsymbol{1}\boldsymbol{g}_1^\top$ from $\boldsymbol{F}_\star$, and then proceeding by analogy with equation 2.14. This leads to

$$\boldsymbol{C}_\Delta = \boldsymbol{B}_\Delta^+ \left(\boldsymbol{F}_\star - \boldsymbol{1}\boldsymbol{g}_1^\top\right) \boldsymbol{G}_\Delta^+,$$

where the matrices $\boldsymbol{B}_\Delta^+$ and $\boldsymbol{G}_\Delta^+$ can be obtained from the QR factorizations of $\boldsymbol{B}_\Delta$ and $\boldsymbol{G}_\Delta^\top$, as before.

2.5.2 Steered Solution. In two (or more) dimensions, it is assumed that the desired filters $G_1(x - t, \theta, \sigma)$ have a common orientation, where $\boldsymbol{v} = (\cos\theta, \sin\theta)^\top$ is the direction of the derivative in 2D. This leads to invariance with respect to translations of the signal in the given direction. The basis filters $G_k(x, \sigma, \theta)$, however, will typically have a range of orientations $\phi_\ell \neq \theta$. This problem can be solved as follows.

Recall from section 2 that the kth-order gaussian derivative is steerable with a basis of size $k + 1$. Now suppose that row k of the matrix $\boldsymbol{G}$ is replaced by $k + 1$ rows, containing sampled filters $G_k(x, \phi_\ell, \sigma)$ at $\ell = 1, \ldots, k + 1$ distinct orientations ϕ_ℓ. The enlarged matrix $\boldsymbol{G}_\phi$ now has dimensions $M_K \times N$, where

$$M_K = \sum_{k=1}^{K} (k + 1) = \frac{1}{2} K(K + 3). \tag{2.17}$$

It follows that there is a $K \times M_K$ steering matrix $\boldsymbol{D}$ such that $\boldsymbol{G} = \boldsymbol{D}\boldsymbol{G}_\phi$ is exactly the $K \times N$ matrix of derivatives at the desired orientation. Moreover, if the approach of section 2.4 is applied to the $M_K \times N$ matrix $\boldsymbol{G}_\phi$, then a solution,

$$\boldsymbol{F} = \boldsymbol{B}\boldsymbol{C}_\phi \boldsymbol{G}_\phi = \boldsymbol{B}\boldsymbol{C}\boldsymbol{G}, \tag{2.18}$$

will be obtained. It follows that the two coefficient matrices are related by $\boldsymbol{C}_\phi = \boldsymbol{C}\boldsymbol{D}$. In summary, if the matrix $\boldsymbol{G}$ contains a sufficient number M_K of differently oriented filters, then a set of translated filters $\boldsymbol{F}$ can be approximated in any common orientation θ. There is no change to the algorithm

described in section 2.4. It should, however, be noted that equation 2.17 shows a trade-off between translation invariance and steerability. Larger translation invariance requires more derivatives, but these become increasingly difficult to steer.

2.5.3 Additive Steered Solution. The steered solution, as described in the previous section, will not automatically be additive, in the sense of equation 2.16. This problem will be solved, with reference to section 2.5.1, by putting an explicitly steered filter $\boldsymbol{g}_\theta^\top$ in place of $\boldsymbol{g}_1^\top$. The first derivative can be steered with respect to a basis of filters at distinct orientations ϕ_1 and ϕ_2 (these would be the first two rows of the $M_K \times N$ matrix $\boldsymbol{G}_\phi$). The standard steering equation (Freeman & Adelson, 1991) can be simplified in this case to

$$\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} \cos\phi_1 & \cos\phi_2 \\ \sin\phi_1 & \sin\phi_2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix},$$

where θ, ϕ_1, and ϕ_2 are known angles. This system can be solved exactly for the unknown coefficients p_1 and p_2, resulting in

$$p_1 = \sin(\phi_2 - \theta)/\delta \quad \text{and} \quad p_2 = \sin(\theta - \phi_1)/\delta$$
$$\text{where} \quad \delta = \sin(\phi_2 - \phi_1). \tag{2.19}$$

It may be noted that if $\phi_1 = 0$ and $\phi_2 = \pi/2$, then the solution reduces to the usual coefficients $p_1 = \cos\theta$ and $p_2 = \sin\theta$ for the construction of the directional derivative from $\mathrm{d}/\mathrm{d}x$ and $\mathrm{d}/\mathrm{d}y$. The additive steered approximation can now be defined, using the new filter $G_1(\boldsymbol{x}, \theta, \sigma)$, as

$$\boldsymbol{F}_\star \approx \boldsymbol{1}\boldsymbol{g}_\theta^\top + \boldsymbol{B}_\Delta \boldsymbol{C}_{\phi\Delta} \boldsymbol{G}_{\phi\Delta} \quad \text{where} \quad \boldsymbol{g}_\theta^\top = p_1 \boldsymbol{g}_{\phi_1}^\top + p_2 \boldsymbol{g}_{\phi_2}^\top. \tag{2.20}$$

This system can be solved in the same way as equation 2.16, which gives a family of steered and shifted filters, as shown in Figure 2. Note that the higher-order filters in $\boldsymbol{G}_{\phi\Delta}$ will be implicitly steered, as described in section 2.5.2.

3 Evaluation

Two issues are addressed in this evaluation, as follows: approximation (the accuracy of the least-squares algorithms from sections 2.4 and 2.5 is established in section 3.1) and characterization (the response of the underlying model from section 1.1 to basic stimuli, as well as to natural images, is analyzed in sections 3.2 and 3.3, respectively). The issues of approximation and characterization are addressed separately in order to avoid mixing different

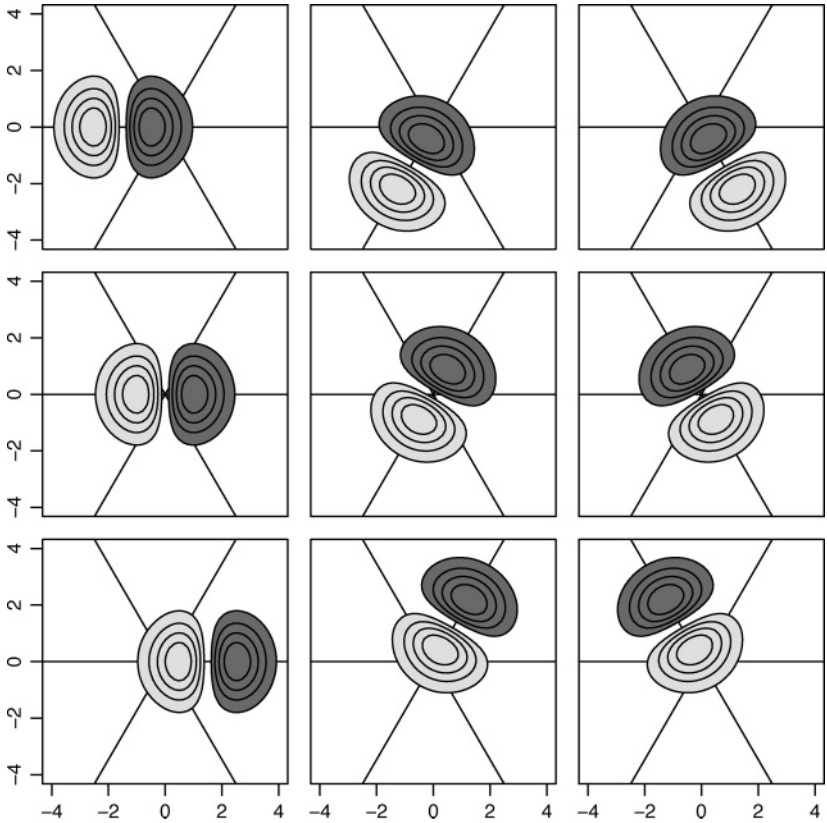

Figure 2: Synthesis of orientation-tuned subunits. The nine filters were synthesized from eight oriented derivatives centered at the origin (with of $\sigma = 1$). The steered additive solution of section 2.5.3 was used. (Middle row) The G_1 filter is steered to the axes of a hexagonal lattice; $\theta = 0^\circ, 60^\circ, 120^\circ$. (Top and bottom rows) Offset filters, synthesized at shifts of $t = \pm\rho$ in the corresponding directions (with $\rho = 1.5$). Note that each column can be interpreted as three subunits of an orientation-tuned complex cell. The filter amplitudes are scaled to the range $[-1, 1]$, with contour lines separated by increments of 0.2 units.

sources of error. Hence, section 3.1 will evaluate the approximate filters F, while sections 3.2 and 3.3 will analyze the ideal filters $F_\star$.

3.1 Approximation Error. The accuracy of the filter approximations will be evaluated in this section, and it will be shown that the least-squares methods are superior to the original Maclaurin expansion. The evaluation is based on the root-mean-square (RMS) error between the target and synthetic filters.

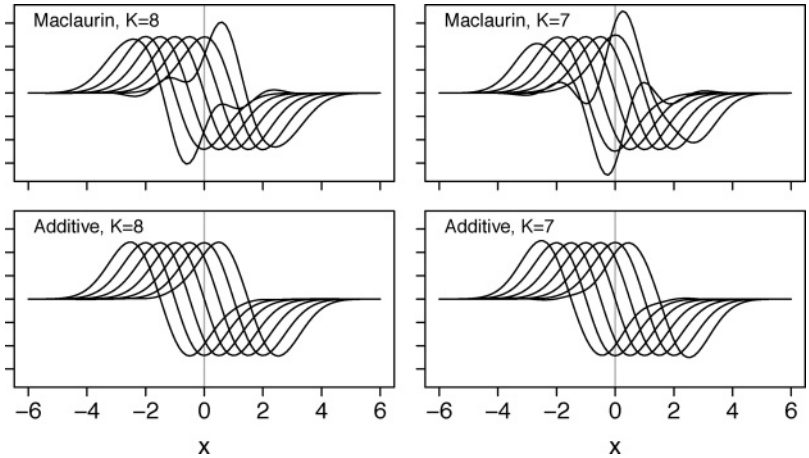

Figure 3: Filter deformation. (Top left) The eighth-order Maclaurin synthesis, equation 2.8, of filters $\sigma = 1$, over a range $\rho = \pm 1.5\sigma$ of offsets. Large errors are visible in the most extreme filters. (Top right) The approximation is much worse if the order of the basis is reduced by 1. (Bottom left, right) The least-squares approximation is much better, even if the additivity constraint is enforced.

The accuracy of a given filter synthesis method is determined by two variables: the range of offsets and the number of available derivatives (size of the basis). Better approximations can, in general, be obtained by reducing the range of offsets or increasing the size of the basis, or both. The range $\rho = 1\sigma$ is the smallest that results in a unimodal impulse response, as will be shown in section 3.2.1. It is therefore important to analyze the corresponding approximation. In addition, the larger range, $\rho = 1.5\sigma$, will be analyzed. This leaves the size of the basis (for which there is no prior preference) to be varied in each case.

The method of evaluation is illustrated in Figure 3. It can be seen that the farthest-offset filters begin to depart from the target shape. The RMS difference between the ideal and approximate filters, for each test, was measured over 51 offsets t_i in the range $\pm\rho$. Each filter was sampled at 101 points x_j in the range $\pm 6\sigma$, which contains the significantly nonzero part of all filters (see Figure 3).

The RMS error, for basis sizes $K = 3, \ldots, 10$ is shown in Figure 4. The meaning of the RMS error, in terms of filter distortion, can be gauged with reference to Figure 3. For example, it can be seen that the Maclaurin model with $K = 8$ and $\rho = 1.5$ is poor, and this corresponds to a point around the middle of the RMS axis in Figure 4.

In the case of the Maclaurin approximation (top row) it can be seen that the error increases rapidly and monotonically with respect to the offset. The

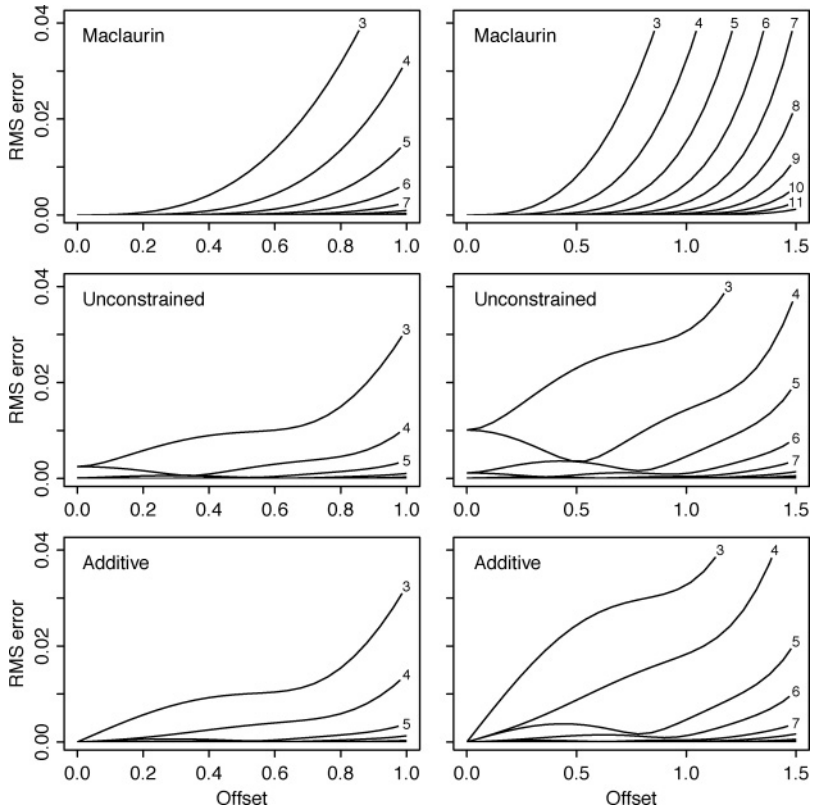

Figure 4: Error versus offset. (Top row) The Maclaurin error rises quickly as the target filter is offset from the center of the basis. Each line represents a different basis size, $k = 3, \ldots, 12$, as indicated. The left and right plots show ranges $\rho = 1\sigma$ and $\rho = 1.5\sigma$, respectively. (Middle row) The unconstrained least-squares approximation is much better, especially for high-order bases. (Bottom row) The additive approximation is also good and ensures that the error is 0 when there is no offset (as in the Maclaurin case).

pattern is more complicated for the least-squares approximations, because the error has been minimized over an interval $\pm\rho$, which effectively truncates the basis functions in x. Nonetheless, the lines corresponding to the different basis sizes remain nested; they cannot cross, because increasing the size of the basis cannot make the approximation worse. It is, however, possible for the lines to meet. In particular, the unconstrained lines meet in pairs at $t = 0$. This is because the target function at 0 offset is anti symmetric. It follows that the incorporation of a symmetric basis function G_{2k} cannot improve an existing approximation of order $2k - 1$. In the case of

the additive approximation, all lines meet at $t = 0$, where the error is 0 by construction.

3.2 Response to Basic Signals. A number of basic signals will be introduced below, and the ideal responses will be derived; further examples are given in Hansard and Horaud (2010). The responses are ideal in the sense that the error of the least-squares approximations (see sections 2.4–2.5.3) will be ignored. This is primarily in order to obtain useful results, but there are two further justifications. First, we demonstrated in the preceding section that the approximations are good over an appropriate range ρ. Second, the approximation error can be made arbitrarily low by using a large enough basis for the given range.

Recall that the offset filters $F(t, x)$ are copies of the gaussian derivative $G_1(x, \sigma)$. It follows from equation 2.6 that the response function is covariant to the shift t in the sense that

$$R(t, u) = R(0, u - t). \tag{3.1}$$

It therefore suffices to obtain the linear response for the case $t = 0$, as the other responses are simply translations of this function. The linear response in this case is $R(0, u) = G_1^1(x, \sigma) \star S(x - u)$, by analogy with equation 2.4. Note that the normalized filter has been used, as defined in equation 2.3. It follows that $R(0, u)$ can be obtained by blurring the signal with the filter $G_0^1(x, \sigma)$ and then differentiating the result. The complex response $C(u)$ is given by the max operation, equation 2.7. Evidently $C(u)$ is the upper envelope of the family $|R(t, u)|$, but it is possible to be more precise than this. In particular, the shift insensitivity of the model can be quantified by determining the intervals of u over which $C(u)$ is constant, as described below.

The response $|R(0, u)|$ to a basic signal $S(x - u)$ can be symmetric or antisymmetric, and periodic or aperiodic. However, a common property of the responses considered here is that the local maxima are all of equal height. Let $|R(0, u^\star)| = R^\star$ be a local maximum, and suppose that u is within range of this maximum, meaning that $|u - u^\star| \leq \rho$. It follows that $C(u) = R^\star$, because the maximum in equation 2.7 will be found at $t = u - u^\star$, and $|R(u - u^\star, u)| = |R(0, u^\star)| = R^\star$ by equation 3.1. An intuitive summary of this is that each local maximum $|R(0, u^\star)|$ generates a plateau $C(u^\star \pm \rho) = R^\star$ in the complex response. In order to make use of this interpretation, the function $V(u)$ will be defined as the signed distance $u - u^\star$ to the nearest local maximum of $|R(0, u)|$. It follows that

$$C(u) = \begin{cases} R^\star & \text{if } |V(u)| \leq \rho \\ \max_{|t| \leq \rho} |R(t, u)| & \text{otherwise} \end{cases}. \tag{3.2}$$

This explicitly identifies the intervals, $|V(u)| \leq \rho$, over which $C(u)$ is constant. Note that if $V(u) > \rho$, then the original definition 2.7 is used. The functions $R(0, u)$ and $V(u)$, as well as the constant $R^\star$, will now be derived for each of the basic signals. It should be emphasized that $V(u)$ and $R^\star$ are used only to characterize the response; they are not part of the computational model.

3.2.1 Impulse. The first test signal to be considered is the unit impulse, which can be used to characterize the initial linear stage of the model. The impulse is defined as

$$S_\sigma(x) = \delta(x), \tag{3.3}$$

where $\delta(x)$ is the Dirac distribution. It follows that the linear response $G_1^1 \star S_\sigma$ is just the original normalized derivative filter,

$$R_\sigma(0, u) = G_1^1(u, \sigma). \tag{3.4}$$

The maxima of the linear response can be found by differentiating $R_\sigma(0, u)$ and setting the result to 0. The derivative contains a factor $\sigma^2 - u^2$, and so the 0's are at $\pm\sigma$. The peak is at $-\sigma$, and it follows that the maximum response is

$$R_\sigma^\star = G_1^1(-\sigma, \sigma). \tag{3.5}$$

Both extrema become peaks in $|R_\sigma(0, u)|$, and the extent of the response plateau is determined by the minimum distance from these. The distance function for the impulse response can now be defined as

$$V_\sigma(u) = u - \mathrm{sgn}_+(u)\,\sigma, \tag{3.6}$$

where sgn_+ is the sign function with the convention $\mathrm{sgn}_+(0) = 1$. If $u = 0$, then $|V_\sigma(u)| = \sigma$, and it follows from equation 3.2 that $C(u) \neq R_\sigma^\star$ if $\sigma > \rho$. It has already been established that $|R_\sigma(0, u)|$ has maxima of $R_\sigma^\star$ at $\pm\sigma$, which implies that the response $C(u)$ will be bimodal unless

$$\sigma \leq \rho, \tag{3.7}$$

as illustrated in Figure 5. This condition is strictly imposed, as it would be undesirable to have a bimodal response to a unimodal signal. In general, ρ should be made as large as possible for a given σ in order to achieve as much shift invariance as possible. Recall, for example, that the least-squares approximations in section 3.1 were demonstrated for $\rho = 1.5\sigma$.

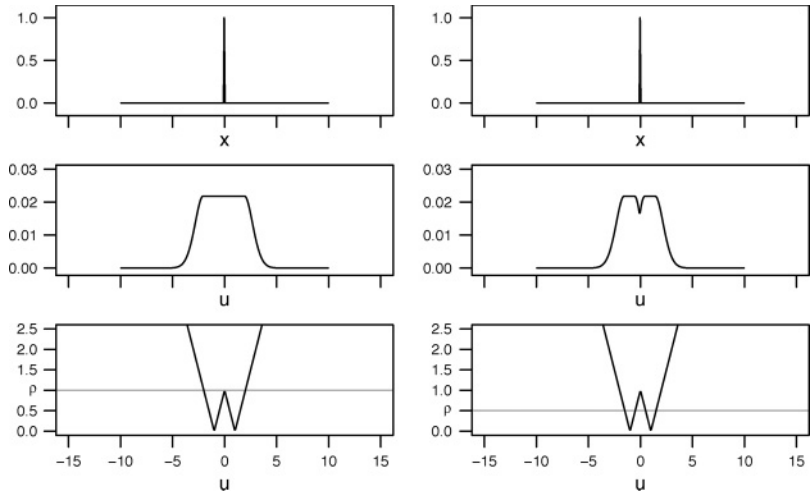

Figure 5: Impulse response. (Left column) The top plot shows the unit impulse, $S_\sigma(x)$, as in equation 3.3. The middle plot shows the response $C(u)$. The bottom plot shows the distance function $|V_\sigma(u)|$, as in equation 3.6, along with the value of the maximum offset $\rho = 1\sigma$. The response is constant, $C(u) = R^\star_\sigma$, when $|V_\sigma(u)| \leq \rho$. The critical case, $\sigma = 1$, $\rho = 1$ is plotted. (Right column) As before, except $\rho = 0.5\sigma$. The response becomes bimodal, which shows the importance of the condition $\sigma \leq \rho$.

3.2.2 Step. The second test signal to be considered is the unit step function. This is arguably the most important example because it is the basic model for a luminance edge. Indeed, the current model is optimized for the detection of step-like edges, owing to the use of the first derivative as the offset filter (Canny, 1986). The step can be defined from the standard sign function, as follows:

$$S_\alpha(x) = \frac{\alpha}{2}\big(1 + \mathrm{sgn}(x)\big). \tag{3.8}$$

The unit step function is related to the integral $\Phi(u, \sigma)$ of the normalized gaussian function $G^0_0(x, \sigma)$ in the following way:

$$\Phi(u, \sigma) = \int_{-\infty}^{u} G^0_0(x, \sigma)\, \mathrm{d}x \tag{3.9}$$

$$= \frac{1/\alpha}{\sigma\sqrt{\pi/2}} \int_{-\infty}^{\infty} G^1_0(x, \sigma) S_\alpha(u - x)\, \mathrm{d}x. \tag{3.10}$$

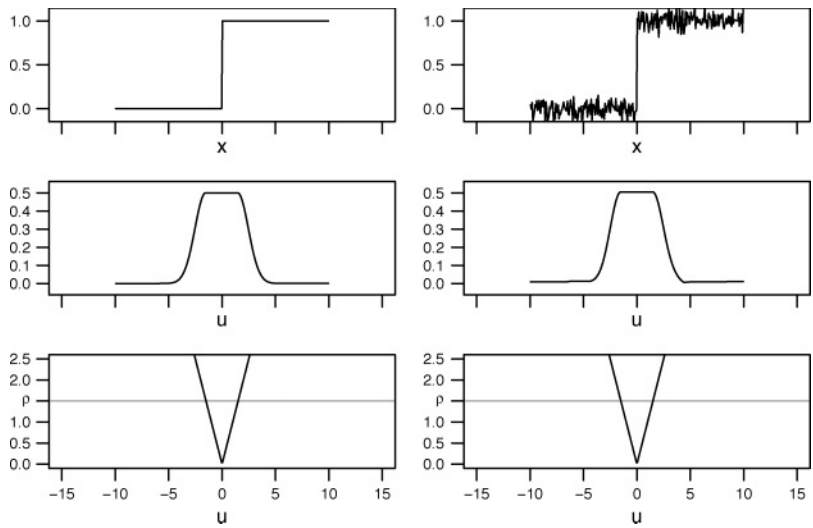

Figure 6: Step response. (Left column) The top plot shows the step edge $S_\alpha(x)$, as defined in equation 3.8. The middle plot shows the response $C(u)$. The bottom plot shows the distance function $V_\alpha(u)$, as in equation 3.13. Note that the response is unconditionally unimodal in this case. (Right column) As before, except with gaussian noise (SD 0.075) added independently at each point. The response $C(u)$ is not significantly affected.

The second integral is the convolution of G_0^1 with S_α, and hence $\Phi(u, \sigma)$ is proportional to the smoothed step edge. The linear response is given by the derivative,

$$R_\alpha(0, u) = \frac{\sigma\sqrt{\pi/2}}{1/\alpha} \frac{\mathrm{d}}{\mathrm{d}u} \Phi(u, \sigma) \tag{3.11}$$

$$= \frac{\alpha}{2} G(u, \sigma). \tag{3.12}$$

This shows that the basic response is simply an unnormalized gaussian, located at the step discontinuity. The maximum response and the signed distance function are evidently

$$R_\alpha^\star = \alpha/2 \quad \text{and} \quad V(u) = u. \tag{3.13}$$

Let $\big(u,\ R(u)\big)$ be the Cartesian coordinates of the response curve. The final response can be constructed from the gaussian, equation 3.12 by inserting the plateau $(\pm\rho, \alpha/2)$ in place of the maximum point $(0, \alpha/2)$. This is illustrated in Figure 6.

3.2.3 Cosine. The third class of signals to be considered are the sines and cosines. These are of central importance owing to their role in the Fourier synthesis of more complicated signals. Furthermore, these functions are used to construct the 2D grating patterns that are commonly used to characterize complex cells. It will be convenient to base the analysis on the cosine function

$$S_\xi(x) = \cos\big(2\pi\xi x\big), \tag{3.14}$$

where ξ is the frequency. The Fourier transforms $g(x) \mapsto \mathcal{F}_x[g](\eta)$ of the filter $G_0^1(x,\sigma)$ and signal $S_\xi(x)$ are

$$\mathcal{F}_x\big[G_0^1\big](\eta) = \sigma\sqrt{\pi/2}\,G\big(\eta, 1/(2\pi\sigma)\big), \tag{3.15}$$

$$\mathcal{F}_x\big[S_\xi\big](\eta) = \tfrac{1}{2}\big(\delta(\eta-\xi)+\delta(\eta+\xi)\big), \tag{3.16}$$

respectively, where η is the frequency variable. The convolution $G_0^1 \star S_\xi$ can be obtained from the inverse Fourier transform of the product $\mathcal{F}_x\big[G_0^1\big]\,\mathcal{F}_x\big[S_\xi\big]$. The resulting cosine is attenuated by a scale factor $\mathcal{F}_x\big[G_0^1\big](\xi)$, because $\mathcal{F}_x\big[S_\xi\big](\eta)$ is 0 unless $|\eta| = \xi$. Differentiating $\cos(2\pi\xi x)$ with respect to x gives $-\sin(2\pi\xi x)$, along with a second scale factor of $2\pi\xi$. The amplitude of the linear response is given by the product of the two scale factors $2\pi\xi$ and $\mathcal{F}_x\big[G_0^1\big](\xi)$, which can be expressed as

$$R_\xi^\star = \sigma\xi\pi^{3/2}\sqrt{2}\,G\big(\xi, 1/(2\pi\sigma)\big).$$

It can be seen that the amplitude depends on the scale σ of the filter, as well as on the frequency ξ of the signal. The complete linear response is given by

$$R_\xi(0,u) = -R_\xi^\star \sin\big(2\pi\xi u\big).$$

Note that a phase shift u_0 can be introduced, if required, by substituting $u - u_0$ for u. The rectified sine $|R_\xi(0,u)|$ is another periodic function, of twice the frequency. The peaks of this function are separated by a distance $\frac{1}{2\xi}$, and so

$$V_\xi(u) = \left(u \bmod \frac{1}{2\xi}\right) - \frac{1}{4\xi} \tag{3.17}$$

is a suitable distance function for the cosine signal, equation 3.14. The case of sine signals is analogous, with sin replaced by cos in the linear response (see section 3.2.3), and u replaced by $u - \frac{1}{4\xi}$ in the distance function, equation 3.17.

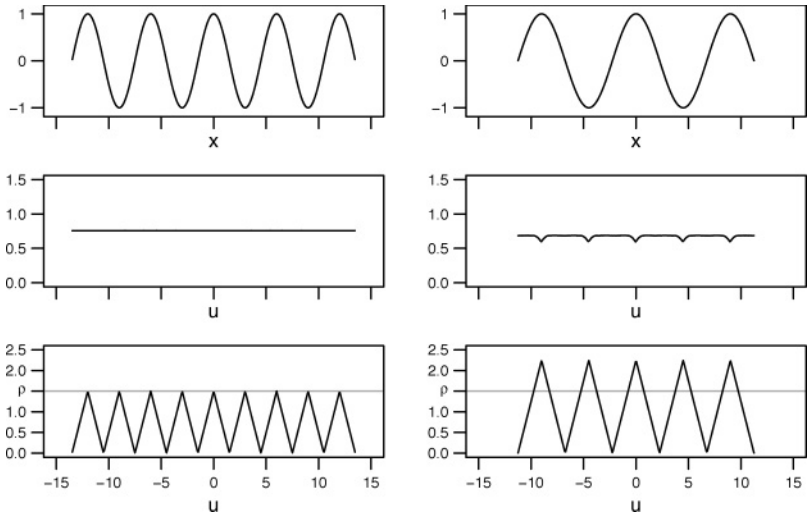

Figure 7: Cosine response. (Left column) The top plot shows the cosine signal $S_\xi(x)$, as in equation 3.14, of frequency $\xi = \frac{1}{6}$. The middle plot shows the response $C(u)$, which is constant. The bottom plot shows the distance function $|V_\xi(u)|$, as in equation 3.17. Note that the critical case is plotted, in which the peaks of $|V_\xi(u)|$ touch the line $\rho = 1.5\sigma$. The response is also constant for any higher frequency. (Right column) As for the left column, but for a lower frequency, $\xi = \frac{1}{9}$. The distance function now crosses the line $\rho = 1.5\sigma$, and corresponding notches appear in $C(u)$.

It is important to see that the system response is entirely constant for frequencies that are not too low. Specifically, the extreme values of equation 3.17, with respect to u, are $\pm\frac{1}{4\xi}$, from which it follows that the response is identically $R^\star_\xi$ if $\xi \geq \frac{1}{4\rho}$. The corresponding constraint on the wavelength $\frac{1}{\xi}$ is

$$1/\xi \leq 4\rho. \tag{3.18}$$

In order to interpret this result, recall that $\rho \geq \sigma$ is required for a unimodal impulse response (see equation 3.7). Furthermore, in section 3.1, it was shown that $\rho \approx 1.5\sigma$ is achievable in practice. This means that a constant response can be expected for frequencies as low as $\xi = \frac{1}{6\sigma}$, as shown in Figure 7.

3.3 Response to Natural Images. This section makes a basic evaluation of the differential model, using the objective function of slow feature analysis (Berkes and Wiskott, 2005; Wiskott and Sejnowski, 2002). The

procedure is as follows. Each 1024×768 gray-scale image is decomposed into $i = 1, \ldots, 36$ orientation channels θ_i at scale $\sigma = 2$ pixels. This corresponds to a set of simple-cell responses $S_1(\boldsymbol{x}, \sigma, \theta_i)$, with an angular separation of 5 degrees. The steerability of S_1 is not used (i.e., a separate convolution is done for each θ_i) in order to minimize any angular bias in the image sampling. A set of straight tracks,

$$\begin{aligned} &\boldsymbol{x}_{ijk} = \boldsymbol{p}_j \pm k\Delta \times (\cos\theta_i,\ \sin\theta_i)^\top, \\ &\text{where} \quad j = 1, \ldots, m \quad \text{and} \quad k = 0, \ldots, n, \end{aligned} \tag{3.19}$$

is sampled from each 2D response. The $m = 100$ random points $\boldsymbol{p}_j$ are sampled from a uniform distribution over the image; the sign $\pm$ is also random. The resolution Δ is set to 1 pixel, and the number of steps along each path is $n = 99$. This gives a total of 100^2 samples from each orientation channel. The responses at nonintegral positions $\boldsymbol{x}_{ijk}$ are obtained by bilinear interpolation. The samples are nonnegative by definition, and a global scale factor γ is used to make the overall ijk-mean of $\gamma S_1(\boldsymbol{x}_{ijk}, \sigma, \theta_i)$ equal to $\frac{1}{2}$. The mean simple-cell response is then computed in each orientation channel,

$$E_S(i) = \frac{1}{mn} \sum_{j=1}^{m} \sum_{k=0}^{n} \gamma S_1\big(\boldsymbol{x}_{ijk}, \sigma, \theta_i\big), \tag{3.20}$$

where the scaling by γ ensures that $\sum_i E_S(i) = \frac{1}{2}$. The mean quadratic variation along the paths is also computed in each orientation channel:

$$Q_S(i) = \frac{1}{mn} \sum_{j=1}^{m} \sum_{k=1}^{n} \Big| \gamma S_1\big(\boldsymbol{x}_{ijk}, \sigma, \theta_i\big) - \gamma S_1\big(\boldsymbol{x}_{ij[k-1]}, \sigma, \theta_i\big) \Big|^2. \tag{3.21}$$

The coordinates $\boldsymbol{x}_{ijk}$ and $\boldsymbol{x}_{ij[k-1]}$ represent adjacent points (separated by Δ) on the jth path in the ith channel. In summary, $E_S(i)$ measures the average response for orientation θ_i, and $Q_S(i)$ measures the average spatial variation of this response in direction θ_i. Slow feature analysis finds filters that minimize the quadratic variation. The orientation tuning and slowness measurements are plotted in Figures 8 and 9 as a function of θ_i by attaching vertical bars $\pm\frac{1}{2}\sqrt{Q_S(i)}$ to each point E_i. Each test is then repeated, using the complex response $C(\boldsymbol{x}, \sigma, \theta)$ in place of the simple response $S_1(\boldsymbol{x}, \sigma, \theta)$, giving measurements $E_C(i)$ and $Q_C(i)$.

Three test images with a dominant global orientation are used. First, a vertical cosine grating, $I_{\cos}(x, y) = \frac{1}{2}(1 + \cos(2\pi\xi x))$. The range is set to $\rho = 1.5\sigma$, as usual, and the wavelength is set to $\frac{1}{\xi} = 8\sigma$. These values do not satisfy the limit, equation 3.18, which ensures that the complex response will not be trivial. The simple-cell response is shown in Figure 8 (top left),

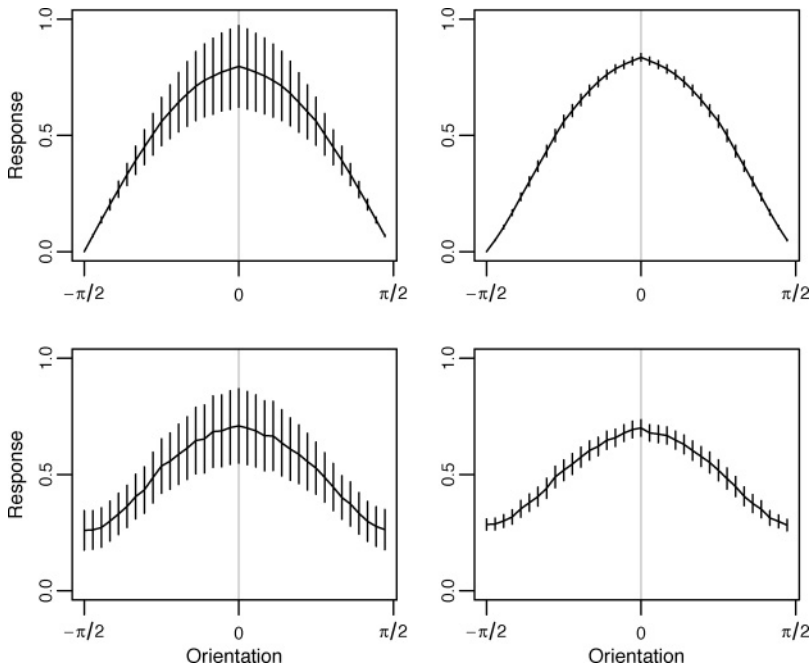

Figure 8: Cosine response. (Top left) Average simple cell response $E_S(i)$ to a vertical cosine grating of wavelength 8σ. The curve indicates the mean response in each of 36 orientation channels θ_i, and has a clear peak at 0. The vertical bars $\pm\frac{1}{2}\sqrt{Q_S(i)}$, indicate the RMS spatial variation of the response in the preferred direction of each orientation channel. (Top right) Complex cell response $E_C(i)$ to the same image. The orientation tuning is preserved, but the variability $\pm\frac{1}{2}\sqrt{Q_C(i)}$ of the response is greatly reduced. (Bottom left, right) As before, but with noise added to the image.

and two effects should be noted. First, the response is tuned to the dominant orientation $\theta = 0$, as can be seen from the unimodal shape of the curve. Second, there is a large variation in the response when the tracks are orthogonal to the grating, as shown by the large bars around $\theta = 0$. This is because the filter falls in and out of phase with the image as it moves horizontally. The corresponding complex response is shown in Figure 8 (top right): the orientation tuning is preserved, while the response variation is greatly reduced. Figure 8 (bottom) repeats the test, but with noise added to the cosine grating, $I = 0.25 \times I_{\cos} + 0.75 \times I_{\text{uni}}$, where each pixel in I_{uni} is independently sampled from the uniform distribution on $[0, 1]$. This means that a variable simple response is obtained as the filter moves in any direction because the image is now truly 2D. The complex response reduces the variation, as shown in Figure 8 (bottom).

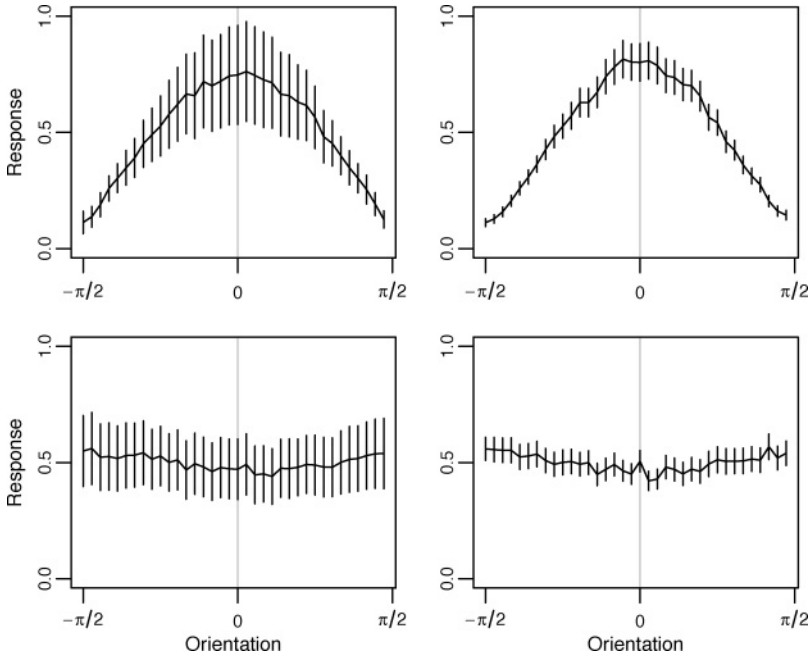

Figure 9: Image response. (Top) As in Figure 8, but using a real image that contains a dominant vertical orientation. Left: The simple response shows variation across all orientation channels $Q_S(i)$. Right: The variation of the complex response $Q_C(i)$ is much lower. (Bottom) As before, but using a natural image, with no dominant orientation.

Figure 9 (top) shows results for a real image which has an orientation structure similar to that of the grating. The results are analogous. Finally, the same test is performed on a natural image, which contains a mixture of foliage and rocks. Figure 9 (bottom) shows that although there is no dominant orientation in the stimulus, the complex response remains much less variable than the simple response.

4 Discussion

It has been shown that a differential model of the complex cell can be constructed from the local jet representation. The differential model, which works naturally in a basis of steerable filters, can be viewed as a constrained version of the Hubel and Wiesel (1962) subunit model. The neural implementation, predictions, and extensions of the model will now be discussed.

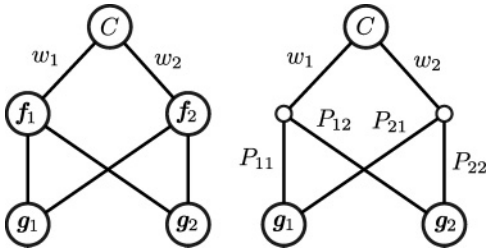

Figure 10: Two neural implementations of the complex cell C. These are schematic representations, with reduced numbers of derivative filters $\boldsymbol{g}_k$ and offset filters $\boldsymbol{f}_i$. (Left) The offset filters are identified with an intermediate layer of simple cells. (Right) The offset filters are implicit in the weighted sums $\boldsymbol{p}_1^\top \boldsymbol{G}$ and $\boldsymbol{p}_2^\top \boldsymbol{G}$ performed by the dendritic tree of C.

4.1 Neural Implementation. The qualitative components in our approach are similar to those of the Gabor energy model. Both models are based on oriented linear filters, which are centered at the same position. Likewise, both models output a combination of the nonlinearly transformed filter responses. The derivative operators $G_k(x, \sigma)$ in the differential model, are interpreted as simple cells, at a common location x (Hawken & Parker, 1987; Young, Lesperance, & Meyer, 2001). These are constructed from LGN inputs, according to the classical model (Hubel & Wiesel, 1962). The offset filters $F(t_i, x)$ have two possible interpretations, shown in Figure 10, as follows.

First, the offset filters $F(t_i, x)$ can be interpreted as an intermediate layer of simple cells, each of which has an RF that is a linear combination of other simple cell RFs. Let the row vector $\boldsymbol{f}_i^\top$ represent $F(t_i, x)$, and let $\boldsymbol{s}$ be the signal vector. It follows that the linear response is

$$r_i = \boldsymbol{f}_i^\top \boldsymbol{s}, \tag{4.1}$$

where $\boldsymbol{f}_i^\top = \boldsymbol{p}_i^\top \boldsymbol{G}$, according to the filter design equation $\boldsymbol{F} = \boldsymbol{P}\boldsymbol{G}$ in equation 2.12. The task of the complex cell C is to compute the (absolute) maximum of the linear responses, r_i. This interpretation corresponds to the HMAX model (Lampl, Ferster, Poggio, & Riesenhuber, 2004; Riesenhuber & Poggio, 1999), but with additional relationships imposed on the underlying simple cells. An advantage of this interpretation is that it predicts a majority of simple cells with few oscillations, as is observed (Ringach, 2002). This follows from the fact that the offset filters are of lower order than their derivatives, together with the fact that an unlimited number of offsets can be obtained from a given derivative basis. Furthermore, suppose that a number of complex cells (e.g., of different orientations) C_n are constructed from the same basis $\boldsymbol{G}$. A new layer of low-order offset filters $\boldsymbol{F}_n$ is

required for each complex cell, and so the high-order filters in $\boldsymbol{G}$ are soon outnumbered in the ensemble $\{\boldsymbol{G}, \boldsymbol{F}_1, \boldsymbol{F}_2, \ldots\}$.

An alternative physiological implementation is as follows. Suppose that the complex cell has $i = 1, \ldots, M$ basal dendrites, each of which branches out to the K simple cells $G_k(x, \sigma)$. The linear response can then be expressed as

$$r_i = \boldsymbol{p}_i^\top(\boldsymbol{G}\boldsymbol{s}), \tag{4.2}$$

where $\boldsymbol{G}\boldsymbol{s}$, the gaussian jet response, is computed first. This interpretation requires no intermediate simple cells. Instead, it places a fixed weight P_{ik} on each dendritic branch and requires a summation to be performed within each dendrite. This seems to be quite compatible with the observed cell morphology; a typical complex cell has a small number of basal dendrites that unlike those of simple cells, are extensively branched (Gilbert & Wiesel, 1979; Kelly & van Essen, 1974).

The dendritic interpretation is more economical in the number of simple cells required, but less compatible with the observed simple-cell statistics (Ringach, 2002). It should be noted that a mixture of the two interpretations in Figure 10 is quite possible. For example, there could be a few intermediate simple cells, with the remaining filters implemented by dendritic summation. In all cases, each complex cell is associated with odd and even simple cells $\boldsymbol{g}_k$, as is observed (Pollen & Ronner, 1981).

The maximum (2.7) over the $|r_i|$ can be approximated by a barycentric combination, $\sum_i w_i |r_i|$. An appropriate vector of weights can be computed as $w_i = |r_i|^\beta / (\alpha + \sum_i |r_i|^\beta)$. This is a form of softmax (Bridle, 1989), with parameters α and β, as detailed in (Hansard & Horaud, 2010). Several neural models of this operation have been proposed (Riesenhuber & Poggio, 1999; Yu, Giese, & Poggio, 2002). For example, the w_i can be interpreted as the outputs of a network of mutually inhibitory neurons, which receive copies of the subunit responses r_i. There is experimental evidence for similar arrangements, with respect to both simple and complex cells (Carandini & Heeger, 1994; Heeger, 1992a; Lampl et al., 2004).

4.2 Experimental Predictions. This section will demonstrate that the response of the new model to broad-band stimuli can easily be distinguished from that of the standard energy model. Some qualitative predictions will also be discussed.

The response of the new model to sinusoidal stimuli is similar to that of the energy model. Both responses are phase invariant, provided that the stimulus frequency is not too low (see Figure 7). Consider, however, a luminance step edge that is flashed (or shifted) across the RF. The Gabor energy is approximately gaussian, as a function of the edge position, with a peak in the center of the RF. The differential response is much flatter,

as shown in Figure 6. This suggests that an empirical measure of kurtosis could be used to distinguish between the two responses, as will be shown below.

Let the odd-symmetric Gabor filter be defined as $F_{\parallel}(x, \xi, \tau) = -G(x, \tau)\sin(2\pi\xi x)$, so that it matches the polarity of G_1. The even filter, following Lehky, Sejnowski, and Desimone (2005), is defined by the numerical Hilbert transform $F_{\perp} = \mathcal{H}(F_{\parallel})$ in order to avoid the nonzero DC component that arises in the cosine-based definition. The Gabor energy of a signal S is then $R_F^2 = (F_{\parallel} \star S)^2 + (F_{\perp} \star S)^2$. The envelope width τ of each Gabor filter is determined from the constraint that the bandwidth be equal to 1.5, which is realistic for complex cells (Daugman, 1985). A self-similar family of Gabor pairs, parameterized by frequency ξ, can now be defined. Quasi-Newton optimization is used to determine a frequency ξ_0, for which $F_{\parallel}(x, \xi_0, \tau) \approx G_1(x, 1)$ in the L^2 sense. Ten Gabor pairs with frequencies $\xi_k = \xi_0 2^{-k\Delta\xi}$ are constructed, where $k = 0, \ldots, 9$. The corresponding differential model, with target filter $G_1(x, \xi_0/\xi_k)$, is also constructed for each pair. Four possible ranges are considered for the differential models by setting $\rho/\sigma = 1.0, 1.25, 1.5, 1.75$.

Let $p_\xi(u)$ be the response distribution, which gives the firing rate for an image edge, as a function of its offset u from the center $u = 0$ of the complex RF. If $P_\xi(u)$ is the cumulative distribution $\int p_\xi(u)\,\mathrm{d}u$, then the (uncentered) kurtosis κ of any $p_\xi(u)$ can be robustly estimated (Crow & Siddiqui, 1967) by

$$K(p_\xi) = \frac{P_\xi^{-1}(1-a) - P_\xi^{-1}(a)}{P_\xi^{-1}(1-b) - P_\xi^{-1}(b)}. \tag{4.3}$$

The values $a = 0.025$ and $b = 0.25$ are used here (making K the ratio of the 95% and 50% confidence intervals). This estimate, which can be computed by linear interpolation between the samples of $P_\xi(u)$, has the following interpretation. Suppose that the offsets u_a and u_b are associated with firing rates a and b, respectively. The response distributions are symmetric, and so the statistic, equation 4.3 is simply the distance ratio $K(p_\xi) = u_a/u_b$, as illustrated in Figure 11. The kurtosis could also be estimated from the empirical moments of $p_\xi(u)$, but equation 4.3 is much less sensitive to noise in the tails of the distribution.

The distributions considered here lie in and around the range from the uniform distribution ($\kappa = \frac{9}{5}, K = 1.9$) to the normal distribution ($\kappa = 3, K \approx 2.91$). It can be seen from Figure 11 that the Gabor response is approximately gaussian, whereas the differential responses are much flatter. Furthermore, note that the line $\rho = \sigma$ in Figure 11 shows the maximum kurtosis of the differential model (determined by equation 3.7), which is still much lower than that of the Gabor energy. Furthermore, it can be seen that the kurtosis is approximately independent of frequency, which simplifies the comparison.

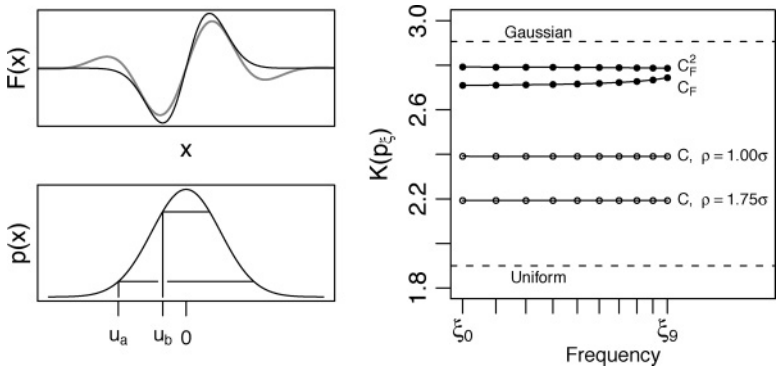

Figure 11: (Top left) Gaussian derivative (black) and Gabor (gray) filters, matched subject to the bandwidth condition. (Bottom left) The kurtosis statistic, equation 4.3, is the ratio of the two horizontal lines, shown here on a gaussian distribution. (Right) Kurtosis of the edge response. Each line represents a complex cell model, parameterized by preferred frequency ξ. The top pair is obtained from the Gabor energy and its square root. The bottom pair delimits the range of possible differential models. The Gabor and differential responses are easily distinguished. The dashed lines are estimates for reference distributions.

It should be noted that a much better fit between G_1 and $F_\parallel$ can be obtained if the bandwidth constraint is relaxed. This, however, makes the energy responses even more kurtotic.

The differential model makes several predictions about the configuration of simple and complex cells. First, like the energy model, it predicts that both odd and even filters are required by the complex cell. Unlike the energy model, it does not require an exact quadrature relationship (indeed, the G_k basis is not orthogonal). More generally, an important property of the differential model is its robustness to deviations from the ideal simple-cell RF profiles. The derivative of gaussian basis was used, in our derivation, for mathematical clarity. However, all that is required is a basis that spans the space of desired subunit filters $F(t_i, x)$.

The differential model also predicts a relationship between the scale σ of the subunits and the radius ρ of the resulting complex receptive field. This prediction, as in the case of the energy model, is probably too strict (i.e., larger complex receptive fields should be possible). However, the complex receptive fields can be extended by allowing multiple scales σ_j in the basis set of the differential model.

A qualitative prediction of our model is that high-order derivative filters are required in order to approximate the target filter over a sufficient range ρ. In particular, we showed in section 3.2.1 that for a unimodal impulse response, $\rho \geq \sigma$ is required. This means, in practice, that derivative filters of

order 5 and beyond must be used in the approximation, as can be seen from Figure 4. This is interesting, because oscillatory filters have been observed in V1 (Young & Lesperance, 2001). These have a natural role as high-frequency processors in the Gabor model. Their role is less clear in the geometric approach, because estimates of the high-order image derivatives are of limited use. Our work suggests that these filters could have a different role in providing a basis for spatially offset filters of low order.

4.3 Future Directions. There are several directions in which this model could be developed. One straightforward extension is to allow filters of different scales (as well as different orders) in the basis set. Preliminary experiments confirm that this extends the range ρ of translation invariance, as would be expected. This means that the complex-cell receptive field could be made larger relative to those of the underlying simple cells. Another extension would be to allow a variety of offset-filter shapes (including even odd symmetric RFs) rather than just the first derivative used here. This would lead to better agreement with the physiological data, which indicates a variety of receptive field shapes among the complex subunits (Gaska, Pollen, & Cavanagh, 1987; Sasaki & Ohzawa, 2007; Touryan, Felsen, & Dan, 2005). It would also be interesting to explore the relationship of our work to the normalization model (Heeger, 1992b; Rust, Schwartz, Movshon, & Simoncelli, 2005), and to other models of motion and spatial processing (Georgeson, May, Freeman, & Hesse, 2007; Johnston, McOwan, & Buxton, 1992).

Our work has concentrated on local shift invariance because this is a defining characteristic of complex cells. However, mechanisms that have other geometric invariances can be constructed in the same scale-space framework. For example, consider the effect of a geometric scaling $(x, y) \to (\alpha x, \alpha y)$ on the operator $G_k^0(x, y, \sigma, \theta)$, which represents the kth derivative of the normalized 2D gaussian, in direction θ. The scaling α has no effect on the shape of the RF, as can be seen from the equation $G_k^0(\alpha x, \alpha y, \alpha\sigma, \theta) = G_k^0(x, y, \sigma, \theta)/\alpha^{2+k}$. This leads to simple relationships between the responses of the RF family $G_k^0(x, y, \sigma_\ell, \theta)$, where $\ell = 1, \ldots, L$ defines a range of scales (Koenderink & van Doorn, 1992; Lindeberg, 1998). Future work will consider more complicated geometric invariances (e.g., 2D affine), in connection with the larger RFs that are found in extrastriate areas.

Another direction would be to consider how the differential model could be learned from natural image data, by analogy with Berkes and Wiskott (2005), Karklin and Lewicki (2009), and Wiskott and Sejnowski (2002). This could be done by fixing the local jet filters (i.e., simple cells) and then optimizing the linear transformation $\boldsymbol{P}$. The transformation could be parameterized by coefficients $\boldsymbol{C}$, given a basis $\boldsymbol{B}$ of smooth functions (e.g., the polynomials that were used here). Alternatively, $\boldsymbol{P}$ could be optimized directly, subject to smoothness constraints on the columns $P_k(t)$. The

variability of the response $C(u)$ would be a suitable objective function for the learning process, as used in slow-feature analysis (Berkes and Wiskott, 2005; Wiskott and Sejnowski, 2002). The combination of geometric and statistical approaches to image analysis is, more generally, a very promising aim.

Acknowledgments

We thank the reviewers for their help with the manuscript.

References

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

Alonso, J.-M., & Martinez, L. M. (1998). Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature Neuroscience, 1*(5), 395–403.

Atherton, T. J. (2002). Energy and phase orientation mechanisms: A computational model. *Spatial Vision, 15*(4), 415–441.

Ben-Shahar, O., & Zucker, S. W. (2004). Geometrical computations explain projection patterns of long range horizontal connections in visual cortex. *Neural Computation, 16*(3), 445–476.

Berkes, P., & Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision, 5*(6), 579–602.

Bridle, J. S. (1989). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fougelman-Soulie & J. Hérault (Eds.), *Neuro-computing: Algorithms, architectures and applictions* (pp. 227–236). Berlin: Springer-Verlag.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*(6), 679–698.

Carandini, M. (2006). What simple & complex cells compute. *J. Physiology, 577*(2), 463–466.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., et al. (2005). Do we know what the early visual system does? *J. Neuroscience, 25*, 10577–10597.

Carandini, M., & Heeger, D. (1994). Summation and division by neurons in primate visual cortex. *Science, 264*, 1333–1336.

Crow, E., & Siddiqui, M. (1967). Robust estimation of location. *Journal of the American Statistical Association, 62*(318), 353–389.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Optical Soc. America, 2*(7), 1160–1169.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 21*, 545–559.

Dobbins, A., Zucker, S. W., & Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature, 329*, 438–441.

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32*(2), 203–218.

Felsberg, M., & Sommer, G. (2001). The monogenic signal. *IEEE Transactions on Signal Processing, 49*(12), 3136–3144.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research, 36*(12), 1839–1857.

Freeman, W. T., & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Trans. PAMI, 13*(9), 891–906.

Gaska, J. P., Pollen, D. A., & Cavanagh, P. (1987). Diversity of complex cell responses to even- and odd-symmetric luminance profiles in the visual cortex of the cat. *Experimental Brain Research, 68*, 249–259.

Georgeson, M. A., May, K. A., Freeman, T.C.A., & Hesse, G. S. (2007). From filters to features: Scalespace analysis of edge and blur coding in human vision. *Journal of Vision, 7*(13), 1–21.

Gilbert, C., & Wiesel, T. (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature, 280*(5718), 120–125.

Hansard, M., & Horaud, R. (2010). Complex cells and the representation of local image-structure (Tech. Rep. 7485). Montbonnot, France: INRIA.

Harris, C., & Stephens, M. (1988). A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference* (pp. 147–151).

Hawken, M. J., & Parker, A. J. (1987). Spatial properties of neurons in the monkey striate cortex. *Proc. R. Soc. Lond. B, B 231*, 251–288.

Heeger, D. J. (1992a). Half-squaring in responses of cat striate cells. *Visual Neuroscience*, 9, 427–443.

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

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiology, 160*, 106–54.

Johnston, A., McOwan, P. W., & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc. R. Soc. Lond. B, 250*, 297–306.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiology, 58*(6), 1233–1258.

Karklin, Y., & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature, 457*, 83–86.

Kelly, J., & van Essen, D. (1974). Cell structure and function in the visual cortex of the cat. *J. Physiology, 238*, 515–547.

Kjaer, T. W., Gawne, T. J., Hertz, J. A., & Richmond, B. J. (1997). Insensitivity of V1 complex cell responses to small shifts in the retinal image of complex patterns. *J. Neurophysiology, 78*, 3187–3197.

Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics, 50*, 363–370.

Koenderink, J. J., & van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biological Cybernetics, 55*, 367–375.

Koenderink, J. J., & van Doorn, A. J. (1990). Receptive field families. *Biological Cybernetics, 63*, 291–297.

Koenderink, J., & van Doorn, A. (1992). Generic neighborhood operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 14*(6), 597–605.

Lampl, I., Ferster, D., Poggio, T., & Riesenhuber, M. (2004). Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J. Neurophysiology, 92*, 2704–2713.

Lehky, S. R., Sejnowski, T. J., & Desimone, R. (2005). Selectivity and sparseness in the responses of striate complex cells. *Vision Research, 45*, 57–73.

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision, 30*(2), 117–154.

Martinez, L. M., & Alonso, J.-M. (2003). Complex receptive fields in primary visual cortex. *Neuroscientist, 9*(5), 317–331.

Mechler, F., Reich, D. S., & Victor, J. D. (2002). Detection and discrimination of relative spatial phase by V1 neurons. *J. Neuroscience, 22*(14), 6129–6157.

Mechler, F., & Ringach, D. L. (2002). On the classification of simple and complex cells. *Vision Research, 42*, 1017–1013.

Morrone, M., & Burr, D. (1988). Feature detection in human vision, A phase dependent energy model. *Proc. R. Soc. Lond. B., 235*, 221–245.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978a). Receptive field organization of complex cells in the cat's striate cortex. *J. Physiology, 283*, 79–99.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiology, 283*, 53–77.

Orban, G. A. (2008). Higher order visual processing in macaque extrastriate cortex. *Physiological Reviews, 88*, 59–89.

Perona, P. (1995). Deformable kernels for early vision. *IEEE Trans. PAMI, 17*(5), 488–499.

Petitot, J. (2003). The neurogeometry of pinwheels as a sub-Riemannian contact structure. *J. Physiol Paris, 97*, 265–309.

Pollen, D. A., & Ronner, S. F. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science, 212*(4501), 1409–1411.

Pollen, A. D., & Ronner, S. F. (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on Systems, Man and Cybernetics, 13*, 907–916.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C* (2nd ed.). Cambridge: Cambridge University Press.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025.

Ringach, D. (2002). Spatial structure & symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiology, 88*, 455–463.

Rust, N., Schwartz, O., Movshon, J., & Simoncelli, E. (2005). Spatiotemporal elements of macaque V1 receptive fields. *Neuron, 46*, 945–956.

Sasaki, K. S., & Ohzawa, I. (2007). Internal spatial organization of receptive fields of complex cells in the early visual cortex. *J. Neurophysiology, 98*, 1194–1212.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H., & Heeger, D. J. (1992). Shiftable multiscale transforms. *IEEE Transactions on Information Theory, 38*(2), 587–607.

Skottun, B. C., Valois, R.L.D., Grosof, D. H., Movshon, J. A., Albrecht, D. G., & Bonds, A. B. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision Research, 31*(7/8), 1079–1086.

Spitzer, H., & Hochstein, S. (1988). Complex cell receptive field models. *Progress in Neurobiology, 31*, 285–309.

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

Wiskott, L., & Sejnowski, T. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation, 14*(4), 715–770.

Wundrich, I. J., von der Malsburg, C., & Würtz, R. P. (2004). Image representation by complex cell responses. *Neural Computation, 16*(4), 2563–2575.

Young, R. A., & Lesperance, R. M. (2001). The gaussian derivative model for spatial-temporal vision: II. Cortical data. *Spatial Vision, 14*(3), 321–389.

Young, R. A., Lesperance, R. M., & Meyer, W. W. (2001). The gaussian derivative model for spatial-temporal vision: I. Cortical model. *Spatial Vision, 14*(3), 261–319.

Yu, A., Giese, M., & Poggio, T. (2002). Biophysiologically plausible implementations of the maximum operation. *Neural Computation, 14*(12), 2857–2881.

Received June 18, 2010; accepted January 28, 2011.