HAL open science

Complex Cells and the Representation of Local Image-Structure

Miles Hansard, Radu Horaud

To cite this version:

Miles Hansard, Radu Horaud. Complex Cells and the Representation of Local Image-Structure. [Research Report] RR-7485, INRIA. 2010. inria-00546779

HAL Id: inria-00546779

https://inria.hal.science/inria-00546779v1

Submitted on 14 Dec 2010

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.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Complex Cells and the Representation of Local Image-Structure

Miles Hansard — Radu Horaud

N° 7485

December 2010

Domaine 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7485--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
GRENOBLE - RHÔNE-ALPES

Complex Cells and the Representation of Local Image-Structure

Miles Hansard, Radu Horaud

Domaine : Perception, cognition, interaction
Équipe-Projet Perception

Rapport de recherche n° 7485 — December 2010 — 41 pages

Abstract: The receptive fields of simple cells in the visual cortex can be understood as linear filters. These filters can be modelled by Gabor functions, or by Gaussian derivatives. Gabor functions can also be combined in an 'energy model' of the complex cell response. This paper 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 2-D 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 relevance of the new model to the cortical image representation is discussed.

Key-words: Biological vision, image representation, signal processing

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Cellules Complexes et la Représentation de la Structure Locale d'une Image

Résumé : Les champs récepteurs des cellules simples du cortex visuel peuvent être inteprétées comme de filtres linéaires. Ces filtres peuvent être modélisés par des fonctions de Gabor, ou par des dérivées gausiennes. Les fonctions de Gabor peuvent également être utilisées dans le cadre d'un 'modèle d'énergie' de la réponse d'une cellule complexe. Cet article propose un modèle alternatif des cellules complexes, basé sur des dérivées gaussiennes. Il est très important de tenir compte de la non-sensibilité de la réponse d'une cellule complexe à des petites translations dans le plan image. Le nouveau modèle utilise une combinaison linéaire des premières dérivées du filtre gaussien, en une seule position, pour approximer la première dérivée en des position adjacentes. La réponse maximale, parmi toutes les positions, fournit un signal qui est insensible aux petits déplacements dans l'image. Ce modèle, contrairement aux approches précédentes, est basée sur la théorie multi-échelle du traitement visuel. En particulier, la cellule complexe est construite sur la base de filtres qui répondent à la structure différentielle 2-D de l'image. Les aspects calculatoires du nouveau modèle sont étudiés en une et deux dimensions, utilisant les propriétés d'invariance rotationnelle des dérivées gaussiennes. La réponse du modèle à des images type, comme des contours ou des grilles, est formellement obtenue. La réponse à des images naturelles est également évaluée, utilisant des mesures statistiques de l'insensibilité aux translations. La pertinence du nouveau modèle par rapport à la représentation corticale d'une image est également discutée.

Mots-clés : Vision biologique, représentation d'une image, traitement du signal

Contents

1 Introduction **4**
1.1 Position Invariance 5
1.2 Phase Invariance 5
1.3 Statistical Models 6
1.4 Present Approach 7
1.5 Overview & Contributions 9

2 Differential Model **9**
2.1 Gaussian Derivatives 9
2.2 Filter Arrays 10
2.3 Functional Representation 11
2.4 Matrix Representation 12
2.5 Unconstrained Estimation 13
2.6 Constrained Estimation 16
2.6.1 Additive Solution 16
2.6.2 Steered Solution 16
2.6.3 Additive Steered Solution 17
2.7 Invariant Response 17

3 Evaluation **19**
3.1 Approximation Error 19
3.2 Response to Basic Signals 20
3.2.1 Impulse 23
3.2.2 Step 23
3.2.3 Cosine 25
3.2.4 Gaussian 27
3.2.5 Bar 27
3.3 Response to Natural Images 28

4 Discussion **32**
4.1 Predictions 33
4.2 Extensions 34

Appendices **34**

A The Local Jet Representation **34**

B The Gabor Energy Model **35**

1 Introduction

It is useful to distinguish between simple and complex cells in the visual cortex, as proposed by Hubel and Wiesel (1962). The original classification is based on four characteristic properties of simple cells, as follows. Firstly, the receptive field has distinct excitatory and inhibitory subregions. Secondly, there is spatial summation within each subregion. Thirdly, there is antagonism between the excitatory and inhibitory subregions. Fourthly, 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 and Ringach, 2002). In particular, it has been argued that the principle characteristic of complex cells is the *phase invariance* of the response (Carandini, 2006; Movshon et al., 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 et al., 1997; Mechler et al., 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 et al., 1982; Movshon et al., 1978a; Skottun et al., 1991). If this ratio is less than one, then 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 and Abbott, 2001), owing to the assumed linearity of the underlying spatial filters (Carandini et al., 1997).

Although the physiology of the complex cell is increasingly well-understood, (as reviewed by Martinez and Alonso, 2003; Spitzer and 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 et al. (1996) in the analysis of binocular processing. It is *invariance* to position or phase that is of interest in the present context. A position-invariant model of the complex cell can be constructed from a set of filters at different positions (see sec. 1.1), as described by Hubel and Wiesel (1962). Alternatively, a phase-invariant model can be constructed from a pair of filters of different phases (see sec. 1.2), as described by Adelson and Bergen (1985). The latter *energy model* has been particularly successful in explaining the physiology of stereopsis (as reviewed by Cumming and DeAngelis, 2001).

The present work is motivated by the difficulty of extending the energy model to more complicated 2-D 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 that are located at the same position. 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 complete model, in this sense, can be seen as a re-formulation of the Hubel and Wiesel (1962) scheme.

The position and phase models will now be discussed in more detail, in sections 1.1 and 1.2, respectively. It is also possible to build *statistical* models of complex cells, by analyzing natural image-data, as discussed in section 1.3. An overview of the new model, and its relationship to previous ideas, will be given in section 1.4.

1.1 Position Invariance

Hubel and Wiesel (1962) suggested that the response of a complex cell might be based on inputs from a group of simple cells. 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 and Hochstein, 1988). This scheme is the basis of the *subunit model*, which was further developed and tested by Movshon et al. (1978a,b). 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 and Poggio, 1999). More complicated functions could be computed by allowing local excitation and inhibition between nearby subunits (Mel et al., 1998; Movshon et al., 1978a).

One attraction of the subunit model is that it is analogous to other hierarchical models of visual processing. These include models of Y-ganglion cells (Hochstein and Shapley, 1976), simple cells (Alonso and Martinez, 1998; Hubel and Wiesel, 1962), hypercomplex cells (Hubel and Wiesel, 1962) and inferotemporal cells (Riesenhuber and Poggio, 1999; Serre et al., 2007). 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. Furthermore, if nonlinearities and local interactions are allowed, then quite general computations could be performed by the resulting network (Riesenhuber and Poggio, 1999). In contrast, the physiological data shows that complex cells are consistently localized in position, orientation and spatial frequency (Ferster and Miller, 2000; De Valois et al., 1982). Likewise, complex cells are relatively invariant to small translations of the image, but not to arbitrary geometric transformations (Hubel and Wiesel, 1962).

1.2 Phase Invariance

If a linear filter is applied to a sinusoidal signal, then the output is another sinusoid, though generally of different amplitude and phase. Now suppose that a pair of filters can be constructed, such that the corresponding outputs differ in phase by $\pi/2$. It follows (from the identity $\sin^2\theta + \cos^2\theta = 1$) that the sum of the squared outputs will be invariant to the phase of the input. This is the basis of the *energy model* which has been used in studies of motion (Adelson and Bergen, 1985; Heeger, 1988), stereopsis (Fleet et al., 1996; Sanger, 1988) and spatial vision (Atherton, 2002; Emerson et al., 1992; Wundrich et al., 2004). Alternatively, a phase-invariant response can be obtained from the dynamics of a suitable network of phase-shifted cells (Chance et al., 1999; Tao et al., 2004).

The application of the energy model to spatial vision is motivated by the observed phase-differences in the receptive fields of adjacent simple cells. Indeed, these receptive fields can be well-represented by odd and even Gabor functions (Daugman, 1985; Jones and Palmer, 1987; Pollen and Ronner, 1983). An energy model can be constructed from these filters, as described in appendix B. This, however, puts strong constraints on the shape of the underlying filters. Strictly, they should form a 'quadrature pair', meaning that they differ in phase by $\pi/2$, but are otherwise identical (Adelson and Bergen, 1985). Although the exact construction of such filters is well-known (via the Hilbert transform) it does not seem biologically plausible. Moreover, the use of approximate quadrature filters, such as Gabor functions, can lead to artefacts in the response (Lehky et al., 2005).

There are two more important problems with the energy model, in the present context. Firstly, 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). Secondly, the quadrature filters that are best-suited to the energy model are not convenient for the general description of 2-D image-structure. The concept of phase itself becomes somewhat complicated in two or more dimensions (Felsberg and Sommer, 2001), and the quadrature representation of more complex image-features, such as edge curvature, is unclear. It must be emphasized that 2-D images contain important structures (e.g. luminance saddle-points) that have no analogue in the 1-D signals to which the energy model is ideally suited. This is less problematic in the case of motion detection, where it often suffices to detect space-time *orientation*. A realistic framework for spatial vision, however, must be capable of representing more complex structures (Ben-Shahar and Zucker, 2004; Dobbins et al., 1987).

1.3 Statistical Models

Three statistical models of the complex cell are described here; slow feature analysis, independent subspace analysis, and patchwise generalization. These methods are complementary to the analytic models of sections 1.1 and 1.2. In particular, the statistical models show that the phase and position models can be obtained from abstract principles of neural coding.

Wiskott and Sejnowski (2002) present a model of the complex cell that is based on *slow feature analysis* (Földiák, 1991). The input is the series of image-patches that are observed as a small aperture moves across a set of natural images. The objective is to learn a set of functions that map the temporally-varying patch intensities to a number of approximately constant output signals. These mappings are constructed from monomial functions of the intensities. It is also necessary to impose temporal constraints on the learned filters (zero mean, unit variance, decorrelation) to make the problem well-posed. It has been shown that slow feature analysis leads to Gabor-like receptive fields, and to energy-like mechanisms (Körding et al., 2004; Wiskott and Sejnowski, 2002). This model of the complex cell has been developed in detail by Berkes and Wiskott (2005), who have shown that many response properties can be predicted in this way, including phase-invariance, orientation tuning, end-inhibition and side-inhibition.

A different statistical approach, *independent subspace analysis*, is described by Hyvärinen and Hoyer (2000, 2001). Here the signal-vector that encodes the local image-patch is projected onto a collection of different feature-subspaces. The response is based on the norms of these projections, which means that it is invariant to any signal-transformation that can be parameterized by the *directions* of the projected vectors. The subspaces themselves can be learned by multidimensional Independent Component Analysis, in which the statistical dependencies between the norms of the projected vectors are minimized.

A third statistical approach has been developed by Karklin and Lewicki (2009), who address the problem of *generalization* across different image-patches that have been sampled from the same part of the scene (e.g. foliage vs. rocks). This is done by parameterizing the complete patch-covariance matrix (of size $n^2 \times n^2$ if each patch has n^2 pixels) by a collection of oblique axes and associated weights. When presented with a particular patch, the model deforms the covariance matrix along the oblique axes, in order to best represent the observation. A system of axes and weights that provides a compact representation is learned from the data. The axes themselves correspond to filters in the original pixel coordinates. Karklin and Lewicki (2009) show that these filters are localized in space, orientation and frequency, and that they can be used to construct a model of the complex cell.

1.4 Present Approach

The classical scale space model of spatial vision includes the natural representation for both 1-D and 2-D image-structure, as well as an exact and rigorous way to parameterize spatial frequency and orientation (Koenderink, 1984). The scale space is a continuum of increasingly blurred images, which are obtained by convolution with a Gaussian filter. This can also be viewed as the solution of the diffusion equation, with the initial conditions supplied by the original image. The spatial derivatives of the blurred image are well-conditioned, and can be obtained directly by convolution with *derivative of Gaussian* filters. The complete local structure of the image can be expressed by suitable combinations of these derivatives (Koenderink and van Doorn, 1987). Spatial frequency is parameterized by the scale dimension, while orientation is parameterized by the *steerability* of the derivative operators (Freeman and Adelson, 1991). The complete set of derivatives, at a given point, represents the *local jet*, as described in appendix A. The Gaussian derivatives are ubiquitous in computer vision (Canny, 1986; Harris and Stephens, 1988), and have also been used in biological models of motion processing (Johnston et al., 1992; Simoncelli and Heeger, 1998) and spatial vision (Georgeson et al., 2007).

The Gaussian derivatives, like the Gabor functions, can be used to model the receptive fields of simple cells (Hawken and Parker, 1987; Young and Lesperance, 2001; Young et al., 2001). A disadvantage of the Gaussian derivatives, however, is that they are not suitable for use in the energy model. This is primarily because the odd and even filters are not matched in spatial frequency. The present work uses Gaussian derivatives to construct a new model of the complex cell, which is functionally comparable to the energy model. Furthermore, the new model inherits the rigour of the scale space model, as described above.

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 modelled by k-th order directional derivatives of the Gaussian kernel, $G_k(\boldsymbol{x}, \sigma, \theta) = (\boldsymbol{v} \cdot \nabla)^k G_k(\boldsymbol{x}, \sigma)$ 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:

$$S_k(\boldsymbol{x}, \sigma, \theta) = G_k(\boldsymbol{x}, \sigma, \theta) \star S(\boldsymbol{x}). \tag{1}$$

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

$$C(\boldsymbol{x}, \sigma, \theta) = \max_t \left|S_1(\boldsymbol{x} + t\boldsymbol{v}, \sigma, \theta)\right|, \quad \text{where} \quad |t| \leq \rho \tag{2}$$

will remain large, unless the shift exceeds the range ρ. Equations (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 paper (formal derivations are given in sec. 2). The complex cell, which inherits the scale and orientation tuning (σ, θ), has a receptive field of radius ρ, centred on position $\boldsymbol{x}$. It can be seen that (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, in section 1.1, that this model is too general. For example, there is no natural limit on the size ρ of the complex receptive field in (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: Firstly, the model becomes inherently local, because the filters G_k that compute the values S_k are now centred at the same point $\boldsymbol{x}$. Secondly, the filters G_k are symmetric or antisymmetric about the point $\boldsymbol{x}$, and therefore resemble the Gabor functions used in the energy model. Thirdly, the values S_k can be obtained from a linear transformation of the K-th order local jet at $\boldsymbol{x}$, and so this scheme is compatible with the scale space theory described above.

To be more specific, it will be shown that the first-order structure in the neighbourhood $\boldsymbol{x} + t\boldsymbol{v}$, as in (2), can be estimated from a linear combination of the directional derivatives S_k,

$$S_1(\boldsymbol{x} + t\boldsymbol{v}, \sigma, \theta) \approx \sum_{k=1}^{K} P_k(t)\, S_k(\boldsymbol{x}, \sigma, \theta) \tag{3}$$

where the functions $P_k(t)$ are fixed polynomials. This approximation will then be substituted into the right-hand side of (2). It will be shown in section 2.3 that the approximation (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 2.4. An optimal (and image-independent) construction of the polynomials $P_k(t)$ will be given in section 2.5. The case in which the derivatives on the right-hand side of (3) are in another direction $\phi \neq \theta$, is treated in section 2.6.

1.5 Overview & Contributions

The model presented in this paper is quite different from the previous approaches, as explained above. The main contribution is a 'differential' model of the complex cell, which is exactly steerable, and which fits naturally into the geometric approach to image analysis (Koenderink and van Doorn, 1987). This shows that it is possible to analyze the local image geometry, and to obtain a shift-invariant response, using a common set of filters.

The body of the paper is organized as follows. The necessary definitions are made in section 2.1, followed by a formal statement of the new approach in sections 2.2–2.3. An algebraic representation of the approach is developed in 2.4. This leads to an optimization problem, which is solved by least-squares methods in 2.5. It is shown in section 2.6 that the new model (unlike Gabor-based models) is exactly steerable. Section 2.7 describes a biologically plausible implementation of the max operation.

The new model is evaluated in section 3. Firstly the accuracy of the least-squares solution is established in section 3.1, by comparison with a series expansion. The exact response of the model to a range of basic stimuli (such as edges and gratings) is analyzed in section 3.2. These results have not been collected elsewhere, although the methods are standard. Some experiments with natural images are described in 3.3, using the cost-function of slow-feature analysis as a criterion. Conclusions and future directions are stated in section 4.

2 Differential Model

The following notation will be used here. Matrices and vectors are written in bold, e.g. $\boldsymbol{M}$, $\boldsymbol{v}$, where $\boldsymbol{M}^\top$ is the transpose, and $\boldsymbol{M}^+$ is the Moore-Penrose inverse (Press et al., 1992). The Kronecker product of matrices is $\boldsymbol{M} \otimes \boldsymbol{N}$. 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 be reviewed in the following section.

2.1 Gaussian Derivatives

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 and van Doorn, 1987). 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 via the spatial derivatives of the corresponding Gaussian. The k-th order derivatives of a 1-D Gaussian, $G_k = \mathrm{d}^k/\mathrm{d}x^k\, G_0$ can be expressed as

$$G_k(x, \sigma) = \left(\frac{-1}{\sigma\sqrt{2}}\right)^k H_k\left(\frac{x}{\sigma\sqrt{2}}\right) G_0(x, \sigma) \tag{4}$$

$$G_0(x, \sigma) = \exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{5}$$

where $G_0(x, \sigma)$ is the original Gaussian, k is a positive integer, and $H_k(x)$ is the k-th Hermite polynomial. The first four Hermite polynomials are, for reference, $H_0(x) = 1$, $H_1(x) = 2x$, $H_2(x) = 4x^2 - 2$, and $H_3(x) = 8x^3 - 12x$. The first

seven Gaussian derivatives are shown in column one of figure 1. It will also be useful to introduce two normalizations of the Gaussian derivatives.

$$G_k^0(x,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, G_k(x,\sigma) \quad \text{and} \quad G_k^1(x,\sigma) = \frac{1}{2}\, G_k(x,\sigma) \tag{6}$$

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 sec. 3.2).

The two-dimensional Gaussian derivative, in direction θ with $\boldsymbol{x} = (x,y)^\top$, will be written $G_k(\boldsymbol{x},\sigma,\theta) = (\boldsymbol{v}\cdot\nabla)^k G_k(\boldsymbol{x},\sigma)$, as in section 1.4. Two special properties of these filters should be noted. Firstly, 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 2-D filter can be obtained from the product of 1-D filters $G_k(x_\theta,\sigma)$ and $G_0(y_\theta,\sigma)$. Secondly, 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 1-D functions (Freeman and 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. Firstly, it is well established that gradient filters can be used to detect edges, as well as more complex image-features (Canny, 1986; Harris and Stephens, 1988). Secondly, G_1 is the first zero-mean filter in the local-jet representation of the image (see appendix A), which is physiologically and mathematically convenient. The extension to higher-order subunits is straightforward, as discussed in section 4.2.

2.2 Filter Arrays

This section will put the system of simple cells, introduced in section 1.4, into a standard signal processing framework. This will be done in 1-D, in order to simplify the notation. The extension to 2-D is straightforward.

The 1-D version of the simple cell response (1) is $S_k(x,\sigma) = G_k(x,\sigma)\star S(x)$. It follows that the shifted first derivative in (2) can be computed from $S_1(x+t,\sigma) = G_1(x,\sigma)\star S(x+t)$. If the signal translates by an amount u, then this becomes

$$\begin{aligned} S_1(x+t-u,\sigma) &= G_1(x,\sigma)\star S(x+t-u) \\ &= -\int_{-\infty}^{\infty} G_1(x-t,\sigma)\, S(x-u)\,\mathrm{d}x. \end{aligned} \tag{7}$$

where the antisymmetry of $G_1(x,\sigma)$ has been used. Note that the result is the negative *correlation* of the signal $S(x-u)$ and the *offset filter* $G_1(x-t,\sigma)$. It is evident that if t could be kept equal to u, then the signal shift would have no effect on the basic response $S_1(x,\sigma)$. It may be noted that the negation in (7) is convenient, because it turns the standard step-edge correlation $\int G_1(x-t,\sigma)\,(1+\mathrm{sgn}(x-u))\,\mathrm{d}x$ into a non-negative function.

2.3 Functional Representation

It was established in the section 2.2 that the response $S_1(x+t,\sigma)$ can be constructed from the offset filters $G_1(x-t,\sigma)$. This means that the desired approximation (3) can be treated as a filter-design problem. The following notation will be adopted for the offset filters:

$$\begin{aligned} F(0,x) &= G_1(x,\sigma) \\ F(t,x) &\approx G_1(x-t,\sigma) \end{aligned} \tag{8}$$

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 which, 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 (7) as

$$R(t,u) = -\int_{-\infty}^{\infty} F(t,x)\, S(x-u)\, \mathrm{d}x. \tag{9}$$

The complex response at $\boldsymbol{x} = (0,0)^\top$, with reference to (2), can now be expressed as a function of the signal translation u;

$$C(u) = \max_t \big|R(t,u)\big| \quad \text{where} \quad |t| \le \rho. \tag{10}$$

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 slight translation of the signal, then $\partial^2 C/\partial u^2 \approx 0$ at u_0.

The approximation problem in (8) 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, as indicated in (8). If image-derivatives up to order K are available, then the approximation is

$$F(t,x) = \sum_{k=0}^{K-1} \frac{(-t)^k}{k!}\, G_{k+1}(x,\sigma) \tag{11}$$

$$= \frac{-G(x,\sigma)}{\sigma\sqrt{2}} \sum_{k=0}^{K-1} \frac{1}{k!} \left(\frac{t}{\sigma\sqrt{2}}\right)^k H_{k+1}\left(\frac{x}{\sigma\sqrt{2}}\right). \tag{12}$$

where (12) follows by substituting the Hermite polynomial expression (4) into (11), and using the fact that $(-1)^k(-1)^{k+1} = -1$ for all k. The Hermite formulation (12) is primarily of theoretical interest. The key observation is that the filters $G_{k+1}(x,\sigma)$ in (11) 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 also be seen from (12) that the estimated function $F(t,x)$ decreases to zero for large $|x|$, as expected. This follows from the exponential tails of the $G(x,\sigma)$ factor.

The definition (11) is usable in practice, as will be shown in section 3.1. There are, however, two difficulties with the scheme described above. Firstly, although $F(t,x)$ is an approximation of $G_1(x-t,\sigma)$, the nature of this approximation is obscure (the *order of contact* is being maximized). Secondly, as

expected, the approximation (11) 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 (11), leading to

$$F(t,x) = \sum_{k=1}^{K} P_k(t)\, G_k(x,\sigma). \tag{13}$$

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 (11). It follows that the polynomials are

$$P_k(t) = \sum_{j=0}^{K-1} t^j c_{jk} \quad \text{where} \quad 1 < k \leq K. \tag{14}$$

The problem has now been altered to that of finding K^2 appropriate coefficients c_{jk}. This will be treated, in sections 2.5 and 2.6, 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 (13), 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 (11) 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 re-sampling the monomial basis functions as t_i^j and, then repeating (13 & 14). 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 bandlimited (Perona, 1995; Simoncelli et al., 1992). The filters $F(t,x)$ are not bandlimited, 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 (9) 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.

2.4 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 2.5. 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 (14). 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 must therefore have the same dimensions, $M \times K$. The monomials are weighted by the $K \times K$ matrix of coefficients $\boldsymbol{C}$, such that

$$\boldsymbol{P} = \boldsymbol{B}\boldsymbol{C} \tag{15}$$

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 (13) that

$$\boldsymbol{F} = \boldsymbol{P}\boldsymbol{G}. \tag{16}$$

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 (9 & 16) as

$$\begin{aligned} \boldsymbol{r} &= -\boldsymbol{F}\boldsymbol{s} \\ &= -\boldsymbol{P}\boldsymbol{G}\boldsymbol{s}. \end{aligned} \tag{17}$$

This clearly shows that the response $\boldsymbol{r}$ is simply a linear transformation $\boldsymbol{P}$ of the K-th 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 $\boldsymbol{x}$.

2.5 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 (17).

Let the $M \times N$ matrix $\boldsymbol{F}_\star$ contain the *true* derivative filters, such that the ij-th element is $G_1(x_j - t_i, \sigma)$. The approximation $\boldsymbol{F} \approx \boldsymbol{F}_\star$ can be expressed, according to (15 & 16), as the product

$$\boldsymbol{F}_\star \approx \boldsymbol{B}\boldsymbol{C}\boldsymbol{G}. \tag{18}$$

Recall that the 'vec' operator stacks the columns of an $M \times N$ matrix into a single $MN \times 1$ vector. If both sides of (18) are stacked in this way, then the identity $\mathrm{vec}(\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}) = (\boldsymbol{Z}^\top \otimes \boldsymbol{X})\,\mathrm{vec}(\boldsymbol{Y})$ can be applied, where '$\otimes$' is the Kronecker product. This leads to the standard least-squares problem $\mathrm{vec}(\boldsymbol{F}_\star) \approx (\boldsymbol{G}^\top \otimes \boldsymbol{B})\,\mathrm{vec}(\boldsymbol{C})$, which can be solved by the Moore-Penrose inverse:

$$\mathrm{vec}(\boldsymbol{C}) = (\boldsymbol{G}^\top \otimes \boldsymbol{B})^+\,\mathrm{vec}(\boldsymbol{F}_\star). \tag{19}$$

This formulation is, however, computationally inefficient. Although the inverse distributes over the Kronecker product, $(\boldsymbol{G}^\top \otimes \boldsymbol{B})^+ = (\boldsymbol{G}^{+\top} \otimes \boldsymbol{B}^+)$, the resulting matrix is of size $K^2 \times MN$, where the number of filter-samples N could be large (especially in higher dimensions). A more practical formulation can be obtained by applying the $\boldsymbol{X}\boldsymbol{Y}\boldsymbol{Z}$ identity to $(\boldsymbol{G}^{+\top} \otimes \boldsymbol{B}^+)\,\mathrm{vec}(\boldsymbol{F}_\star)$, which gives

$$\boldsymbol{C} = \boldsymbol{B}^+ \boldsymbol{F}_\star \boldsymbol{G}^+. \tag{20}$$

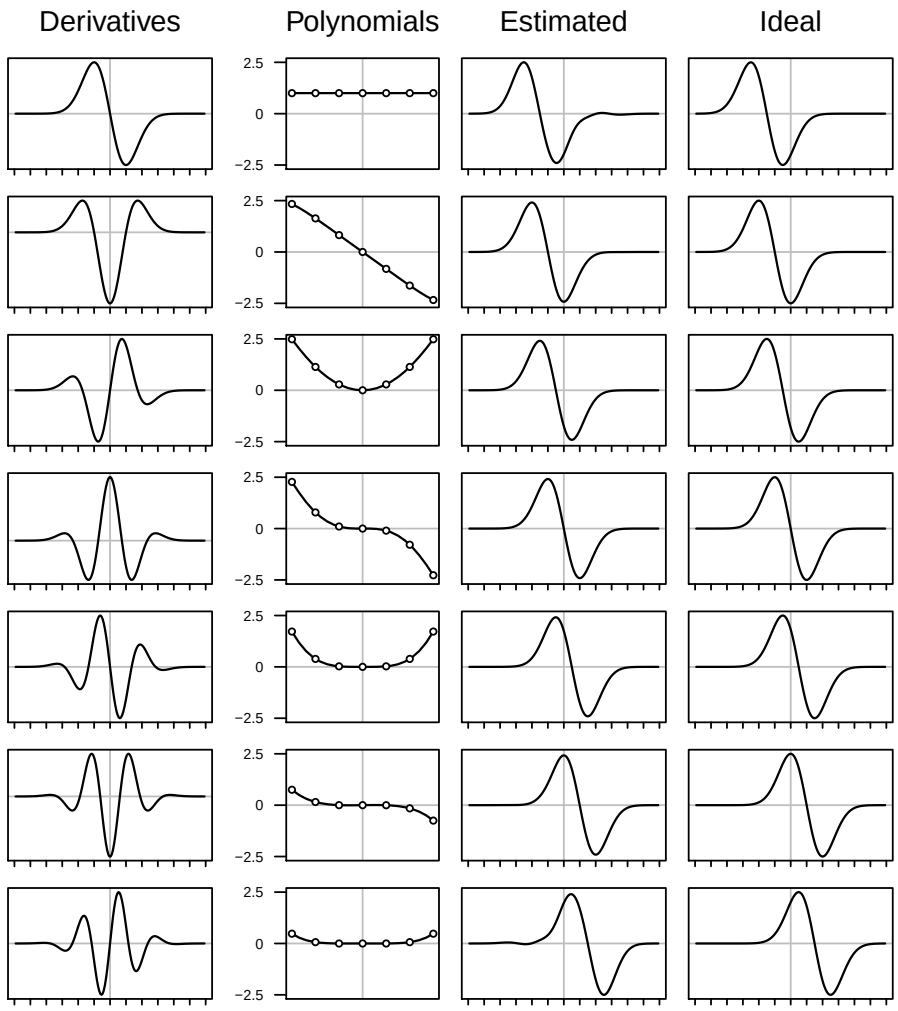

Figure 1: Construction of offset filters. **Column 1:** The Gaussian derivatives $G_k(x, \sigma)$, scaled for display, of orders $1, \ldots, K$, where $K = 7$. **Column 2:** The corresponding polynomial interpolation functions $P_k(t)$, of order $K - 1$. Note that $P_k(t)$ resembles the monomial $t^{(k-1)}$. **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 j-th dot on the k-th polynomial in column two.

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

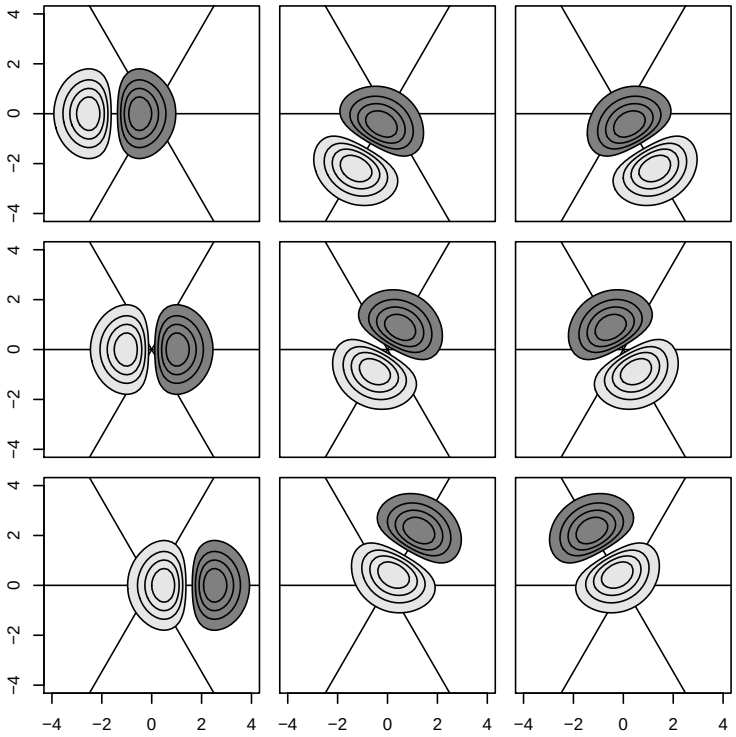

Figure 2: Synthesis of orientation-tuned subunits. The nine filters were synthesized from eight oriented derivatives centred at the origin (with of $\sigma = 1$). The steered-additive solution of section 2.6.3 was used. **Middle row:** The G_1 filter is steered to the axes of a hexagonal lattice; $\theta = 0°, 60°, 120°$. **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.

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 (20) can be obtained via

$$\boldsymbol{B}^+ = \boldsymbol{R}_B^{-1} \boldsymbol{Q}_B^\top \quad \text{and} \quad \boldsymbol{G}^+ = \boldsymbol{Q}_G \boldsymbol{R}_G^{-\top}. \tag{21}$$

The $M \times K$ matrix $\boldsymbol{B}$ depends on the variable number M of offsets, and so the behaviour of (21) with respect to M is important. If the standard monomial basis is used, then $\boldsymbol{B}$ is the well-known *Vandermonde* matrix. This construction is potentially unstable, and so it might be better to use a different basis (Press et al., 1992). In practice however, no difficulties were encountered with either the SVD or QR algorithms described above. For example, the monomial matrix $\boldsymbol{B}$ that was used in the construction of figure 1 had dimensions 21×7, and reciprocal condition number of 0.012, which is well within the acceptable range (Press et al., 1992).

2.6 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. Firstly, if one of the shifts t_i is zero, 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 (8). The second shortcoming of the method in 2.5 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.

2.6.1 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 zero when $t = 0$. This implies the following partitioning of the derivative, monomial and coefficient matrices:

$$\boldsymbol{G} = \begin{pmatrix} \boldsymbol{g}_1 \\ \boldsymbol{G}_\Delta \end{pmatrix} \quad \boldsymbol{B} = \begin{pmatrix} \boldsymbol{1} & \boldsymbol{B}_\Delta \end{pmatrix} \quad \boldsymbol{C} = \begin{pmatrix} 1 & \boldsymbol{0}^\top \\ \boldsymbol{0} & \boldsymbol{C}_\Delta \end{pmatrix} \tag{22}$$

where $\boldsymbol{1}$ is the column-vector of M ones, and $\boldsymbol{0}$ is the column-vector of $(K-1)$ zeros. The $1 \times N$ vector $\boldsymbol{g}_1$ 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 (15), 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 (16), gives the additive approximation

$$\boldsymbol{F}_\star \approx \boldsymbol{1}\boldsymbol{g}_1 + \boldsymbol{B}_\Delta \boldsymbol{C}_\Delta \boldsymbol{G}_\Delta \tag{23}$$

where $\boldsymbol{1}\boldsymbol{g}_1$ is the rank-one matrix containing M identical rows $\boldsymbol{g}_1$. Note that if the i-th row of $\boldsymbol{F}_\star$ corresponds to $t = 0$, then the i-th row of $\boldsymbol{B}_\Delta \boldsymbol{C}_\Delta$ must be zero, this being the evaluation of the polynomials $\sum_{j=1}^{K-1} t^j c_{jk}$ at $t = 0$. It follows that the i-th row of $\boldsymbol{F}_\star$ is exactly recovered from (23) as $\boldsymbol{g}_1$, and so the constraint $F(0,x) = G_1(x,\sigma)$ has been imposed.

The unknown coefficients $\boldsymbol{C}_\Delta$ are recovered by subtracting $\boldsymbol{1}\boldsymbol{g}_1$ from $\boldsymbol{F}_\star$, and then proceeding by analogy with (20). This leads to

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

where the matrices $\boldsymbol{B}_\Delta^+$ and $\boldsymbol{G}_\Delta^{+\top}$ can be obtained from the QR factorizations of $\boldsymbol{B}_\Delta$ and $\boldsymbol{G}_\Delta$, by analogy with (21).

2.6.2 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 2-D. 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 2.1 that the k-th 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

$$M_K = \sum_{k=1}^{K} (k+1) = \tfrac{1}{2} K(K+3). \tag{24}$$

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 2.5 is applied to the $M_K \times N$ matrix $\boldsymbol{G}_\phi$, then a solution

$$\boldsymbol{F} = \boldsymbol{B}\boldsymbol{C}_\phi \boldsymbol{G}_\phi = \boldsymbol{B}\boldsymbol{C}\boldsymbol{G} \tag{25}$$

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 2.5.

2.6.3 Additive Steered Solution

The steered solution, as described in the previous section, will not automatically be additive, in the sense of (23). This problem will be solved, with reference to section 2.6.1, by putting an explicitly steered filter $\boldsymbol{g}_\theta$ in place of $\boldsymbol{g}_1$. 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 and 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

$$\begin{gathered} p_1 = \sin(\phi_2 - \theta)/\delta \quad \text{and} \quad p_2 = \sin(\theta - \phi_1)/\delta \\ \text{where} \quad \delta = \sin(\phi_2 - \phi_1). \end{gathered} \tag{26}$$

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

$$\boldsymbol{F}_\star \approx \boldsymbol{1}\boldsymbol{g}_\theta + \boldsymbol{B}_\Delta \boldsymbol{C}_{\phi\Delta} \boldsymbol{G}_{\phi\Delta} \quad \text{where} \quad \boldsymbol{g}_\theta = p_1 \boldsymbol{g}_{\phi_1} + p_2 \boldsymbol{g}_{\phi_2}. \tag{27}$$

This system can be solved in the same way as (23). Note that the higher-order filters in $\boldsymbol{G}_{\phi\Delta}$ will be implicitly steered, as described in section 2.6.2.

2.7 Invariant Response

The spatial filters $F(t_i, x)$, at M different offsets t_i, have now been constructed. The additive solution (23) is recommended for 1-D filters, while the additive

steered solution (27) is recommended in 2-D. Each of the M filters is convolved with the translating stimulus $S(x-u)$, which results in M linear response-functions $R(t_i, u)$, as in (9). A single response, that is approximately invariant to the shift u, will now be obtained by computing the maximum over t_i, as in (10). This will be done using the neurally plausible 'softmax' function (Bridle, 1989), which does not involve any search procedure. The same approach is used by Riesenhuber and Poggio (1999) in the 'HMAX' model of inferotemporal cells.

The absolute responses $|R(t_i, u)|$ are first mapped to weights $W_\beta(t_i, u) \in [0, 1]$, via a normalized exponential:

$$W_\beta(t_i, u) = \frac{\exp\bigl(\beta|R(t_i, u)|\bigr)}{\sum_{i=1}^{M} \exp\bigl(\beta|R(t_i, u)|\bigr)}. \tag{28}$$

This mapping, which is monotonic, is controlled by the parameter $\beta \geq 0$. If $\beta = 0$ then $W_\beta(t_i, u) = 1/M$, and the mean response, over all offsets, is obtained. More importantly, if $\beta \to \infty$, then $W_\beta = 1$ for the largest of the $R(t_i, u)$, and $W_\beta = 0$ for the others. It follows that the maximum absolute value of $R(t_i, u)$, over all offsets t_i, can be approximated by the barycentric combination

$$C(u) = \sum_{i}^{M} W_\beta(t_i, u)\, |R(t_i, u)| \tag{29}$$

$$\approx \max_i |R(t_i, u)|. \tag{30}$$

It is important to determine a finite but sufficiently large value for β. It will be shown below that this parameter is related to the shape of the response function $R(0, u)$.

The appropriate range for β will now be derived, as determined by the bound that is stated by Yuille and Geiger (2003). Suppose, in general, that the M softmax inputs are

$$\mathcal{Y} = \{x,\, y_1, \ldots, y_{M-1}\}, \quad \text{where } x > y_i \text{ for all } i. \tag{31}$$

The softmax over $\mathcal{Y}$ will be the true maximum if the weight associated with x is 1, while the weights of the y_i are zero. Hence, to find a good approximation, choose a small value ϵ, and require that the x weight be at least $1-\epsilon$. The 'worst case' is that *all* of the y_i are equal to the same value y, meaning that $M-1$ inputs *could* be almost as large as x. The worst case x and y weights, at the ϵ-bound, are therefore

$$1 - \epsilon = \exp(\beta x)/\Sigma \quad \text{and} \quad \epsilon = (M-1)\exp(\beta y)/\Sigma \tag{32}$$

respectively. Note that the y_i weights are all equal to ϵ, and that the denominator is $\Sigma = \exp(\beta x) + (M-1)\exp(\beta y)$, by analogy with (28). Now consider the ratio of the two expressions in (32),

$$\frac{1-\epsilon}{\epsilon} = \frac{\exp(\beta x)}{(M-1)\exp(\beta y)}, \tag{33}$$

and let $\delta = x - y$ be the difference between the largest and second largest inputs (even if the y_i are unequal). Equation (33) can be solved for β, leading to the conclusion

$$\beta \geq \beta_0 \quad \text{where} \quad \beta_0 = \frac{1}{\delta} \log \frac{(M-1)(1-\epsilon)}{\epsilon} \tag{34}$$

as in (Yuille and Geiger, 2003). If $\beta = \beta_0$, then the true maximum is weighted by $1 - \delta$ in the worst case, and by more in general. The number of filters M, and the tolerance ϵ in (34) are known. The response increment δ can also be determined, in the following way.

Suppose, without loss of generality, that the signal shift u is zero, and that the function $R(t, 0)$ has been determined from a given model of the input signal. If $R(t, 0)$ is symmetric and non-increasing around the local maximum $R(0, 0)$, then the difference between the largest and second-largest responses in (34) is $\delta = R(0, 0) - R(\pm\Delta t, 0)$, where $\Delta t = 2\rho/(M - 1)$ is the filter-separation. This means that the appropriate choice of softmax parameter β in (28) is effectively determined by the curvature of the sampled response $R(t_i, 0)$ around the maximum. It will be shown in section 3.2 that simple expressions for $R(t, u)$ can be obtained for several idealized stimuli. This means that, for any tolerance ϵ, a suitable parameter β can be explicitly computed from (34).

It should be noted that any monotonically increasing function of $|R(t, u)|$ could be used in the present model. In particular, $|R(t, u)|^2$ could be used. The square would arguably be closer to the physiological data (Heeger, 1992), and would have the additional benefit of making the response differentiable, as well as non-negative.

3 Evaluation

Two issues are addressed in this evaluation, as follows. *Approximation*: The accuracy of the least-squares algorithms from sections 2.5 and 2.6 is established in section 3.1. *Characterization*: The response of the underlying model from section 1.4 to basic stimuli, as well as to natural images, is analyzed in sections 3.2 and 3.3 respectively. Note that the issues of approximation and characterization are addressed separately, in order to avoid mixing different 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 error (RMSE) between the target and synthetic filters.

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 and/or increasing the size of the basis. 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 approximations over this range. 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 furthest-offset filters begin to depart from the target shape. The RMSE 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 fig. 3). The absolute size of the RMSE obviously depends on this choice of x-sampling. Hence a normalized error was computed, by dividing the least-squares RMSE by the corresponding Maclaurin RMSE, using the same sampling in each case. The results are given in tables 1 and 2, for ranges of $\rho = 1.0\sigma$ and $\rho = 1.5\sigma$, respectively. The results are reported over a range of realistic basis-sizes, from $K = 4$ to $K = 10$.

Each number in the tables 1 and 2 involves a sum over offsets t_i and samples x_j, as described above. It is also interesting to sum over x_j only, in order to reveal the quality of the approximation with respect to the offset. This measure is plotted 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 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 zero 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 zero by construction.

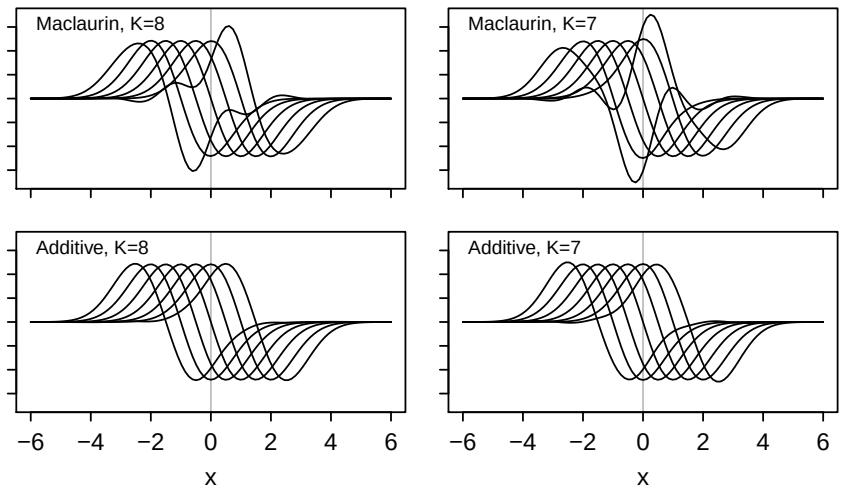

Figure 3: Filter deformation. **Top left:** The eighth-order Maclaurin synthesis (11) of filters $\sigma = 1$, over a range $\rho = \pm 1.5\sigma$ of offsets. Large errors are visible in the most extreme filters. **Top right:** The approximation is much worse if the order of the basis is reduced by one. **Bottom left, right:** The least-squares approximation is much better, even if the additivity constraint is enforced.

3.2 Response to Basic Signals

A number of basic signals will be introduced below, and the ideal responses will be derived. The responses are 'ideal' in the sense that the error of the

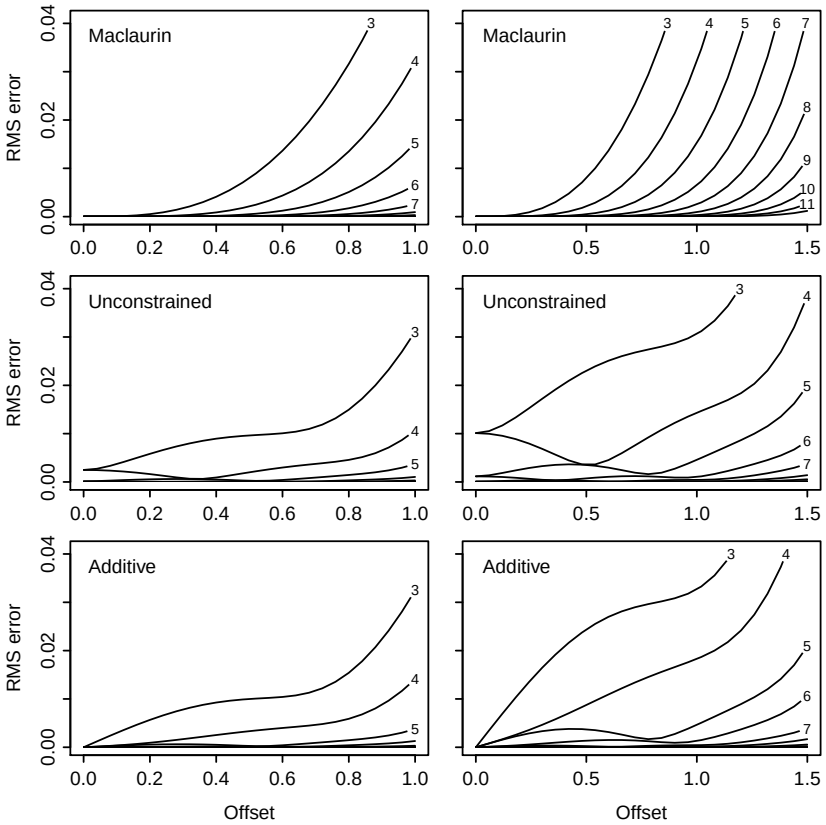

Figure 4: Error vs. offset. **Top row:** The Maclaurin error rises quickly as the target filter is offset from the centre of the basis. Each line represents a different basis-size, $k = 3, \ldots, 12$, as indicated. The left and right plots show ranges $\rho = 1\sigma$, and $\rho = 1.5\sigma$, respectively. **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 zero when there is no offset (as in the Maclaurin case).

Order	4	5	6	7	8	9	10
$100 \times$ RMSE MAC	1.170	0.509	0.203	0.076	0.027	0.009	0.003
UNC/MAC	0.352	0.257	0.168	0.126	0.085	0.064	0.044
ADD/MAC	0.461	0.262	0.207	0.127	0.100	0.064	0.050

Table 1: The first row gives the RMS error of the Maclaurin approximation, of orders 4–10, over the range $\rho = \pm 1.0\sigma$. The second and third rows give the *relative* RMS errors of the unconstrained and additive approximations, respectively.

least-squares approximations (sec. 2.5–2.6.3) will be ignored. This is primarily in order to obtain useful results, but there are two further justifications. Firstly, it has been demonstrated in the preceding section that the approximations are good, over an appropriate range ρ. Secondly, the approximation error can be

ORDER	4	5	6	7	8	9	10
$100 \times$ RMSE MAC	5.511	3.617	2.184	1.226	0.646	0.322	0.152
UNC/MAC	0.288	0.204	0.132	0.097	0.065	0.049	0.033
ADD/MAC	0.384	0.213	0.165	0.100	0.078	0.049	0.038

Table 2: The first row gives the RMS error of the Maclaurin approximation, of orders 4–10, over the range $\rho = \pm 1.5\sigma$. The second and third rows give the *relative* RMS errors of the unconstrained and additive approximations, respectively.

made arbitrarily low, by using a large enough basis for the given range. The derivations are accompanied by *numerical* computations, which are used to verify the analysis.

Recall that the offset filters $F(t, x)$ are copies of the Gaussian derivative $G_1(x, \sigma)$. It follows from (9) that the response function is covariant to the shift t, in the sense that

$$R(t, u) = R(0, u - t). \tag{35}$$

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 (7). Note that the normalized filter has been used, as defined in (6). 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 (10), which is implemented as described in section 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 either symmetric or antisymmetric, and either 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 (10) will be found at $t = u - u^\star$, and $|R(u - u^\star, u)| = |R(0, u^\star)| = R^\star$ by (35). 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

$$C(u) = \begin{cases} R^\star & \text{if } |V(u)| \leq \rho \\ \max\limits_{|t| \leq \rho} |R(t, u)| & \text{otherwise.} \end{cases} \tag{36}$$

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 (10) 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 only used 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

$$S_\sigma(x) = \delta(x) \tag{37}$$

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,

$$R_\sigma(0, u) = G_1^1(u, \sigma). \tag{38}$$

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

$$R_\sigma^\star = G_1^1(-\sigma, \sigma). \tag{39}$$

Both extrema become peaks in $|R_\sigma(0, u)|$, and the extent of the response plateau is determined by the minimum distance from these. It will be convenient to define

$$\mathrm{sgn}_+(u) = \begin{cases} 1 & \text{if } u = 0 \\ \mathrm{sgn}(u) & \text{otherwise} \end{cases} \tag{40}$$

which is the right-continuous sign function. The distance function for the impulse response can now be defined as

$$V_\sigma(u) = u - \mathrm{sgn}_+(u)\,\sigma. \tag{41}$$

If $u = 0$ then $|V_\sigma(u)| = \sigma$, and it follows from (36) 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

$$\sigma \leq \rho \tag{42}$$

as illustrated 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 $\rho = 1.5\sigma$.

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

$$S_\alpha(x) = \frac{\alpha}{2}\big(1 + \mathrm{sgn}(x)\big) \tag{43}$$

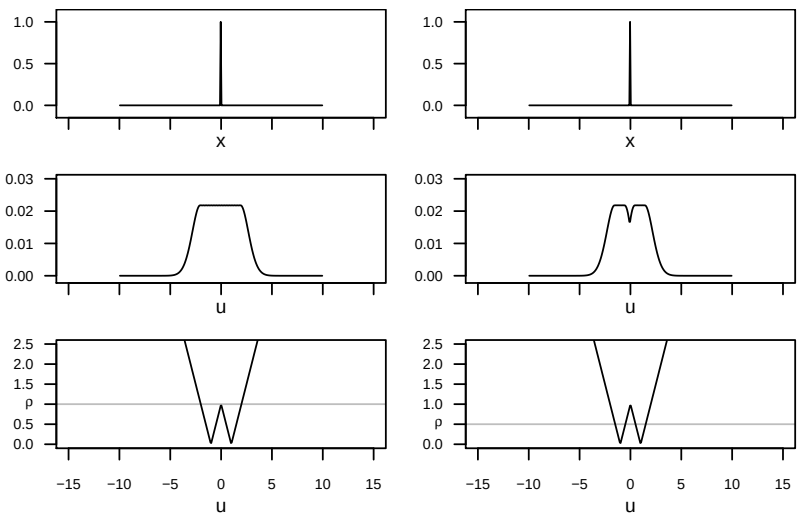

Figure 5: Impulse response. **Left column:** the top plot shows the unit impulse, $S_\sigma(x)$, as in (37). The middle plot shows the response $C(u)$. The bottom plot shows the distance function $|V_\sigma(u)|$, as in (41), along with the value of the maximum offset $\rho = 1\sigma$. The response is 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 $\sigma \leq \rho$.

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:

$$\Phi(u,\sigma) = \int_{-\infty}^{u} G^0_0(x,\sigma)\ \mathrm{d}x \tag{44}$$

$$= \frac{1}{\sigma\sqrt{\pi/2}} \int_{-\infty}^{u} G^1_0(x,\sigma)\ \mathrm{d}x \tag{45}$$

$$= \frac{1/\alpha}{\sigma\sqrt{\pi/2}} \int_{-\infty}^{\infty} G^1_0(x,\sigma)\, S_\alpha(u-x)\ \mathrm{d}x \tag{46}$$

The third integral is the convolution of G^1_0 with S_α, and hence $\Phi(u,\sigma)$ is proportional to the smoothed step-edge. The linear response is given by the derivative,

$$R_\alpha(0,u) = \frac{\sigma\sqrt{\pi/2}}{1/\alpha}\,\frac{\mathrm{d}}{\mathrm{d}u}\,\Phi(u,\sigma) \tag{47}$$

$$= \frac{\sigma\sqrt{\pi/2}}{1/\alpha}\, G^0_0(u,\sigma) \tag{48}$$

$$= \frac{\alpha}{2}\, G(u,\sigma) \tag{49}$$

This shows that the basic response is simply an un-normalized Gaussian, located at the step-discontinuity. The maximum response and the signed-distance func-

tion are evidently

$$R^{\star}_{\alpha} = \frac{\alpha}{2} \quad \text{and} \quad V(u) = u. \tag{50}$$

Let $\big(u, R(u)\big)$ be the Cartesian coordinates of the response curve. The final response can be constructed from the Gaussian (49) by inserting the plateau $(\pm\rho, \alpha/2)$ in place of the maximum point $(0, \alpha/2)$. This is illustrated in in figure 6.

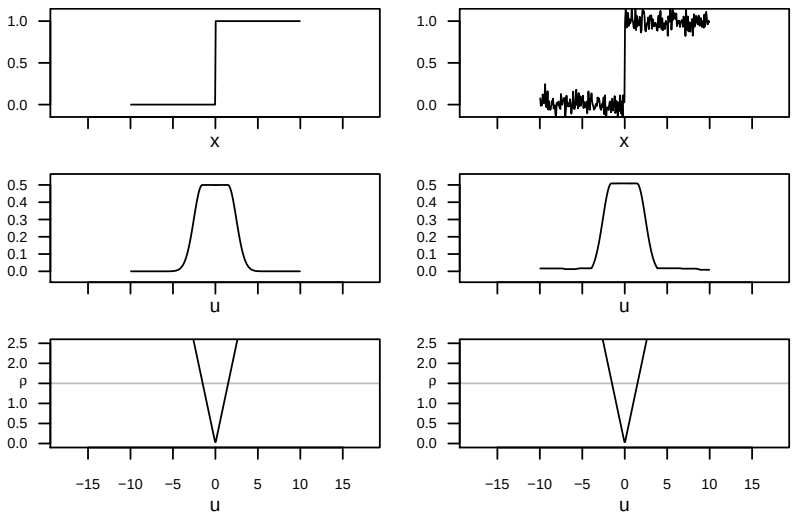

Figure 6: Step response. **Left column:** The top plot shows the step-edge $S_{\alpha}(x)$, as defined in (43). The middle plot shows the response $C(u)$. The bottom plot shows the distance function $V_{\alpha}(u)$, as in (50). Note that the response is unconditionally unimodal in this case. **Right column:** As before, except with Gaussian noise (SD 0.075) added independently at each point. The response $C(u)$ is not significantly affected.

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 2-D grating patterns that are commonly used to characterize complex cells. It will be convenient to base the analysis on the cosine function

$$S_{\xi}(x) = \cos\big(2\pi\xi x\big) \tag{51}$$

where ξ is the frequency. The Fourier transforms $g(x) \mapsto \mathcal{F}_x[g](\eta)$ of the filter $G^1_0(x, \sigma)$ and signal $S_{\xi}(x)$ are

$$\mathcal{F}_x\big[G^1_0\big](\eta) = \sigma\sqrt{\pi/2}\, G\big(\eta, 1/(2\pi\sigma)\big) \tag{52}$$

$$\mathcal{F}_x\big[S_{\xi}\big](\eta) = \tfrac{1}{2}\big(\delta(\eta - \xi) + \delta(\eta + \xi)\big) \tag{53}$$

respectively, where η is the frequency variable. The convolution $G^1_0 \star S_{\xi}$ can be obtained from the inverse Fourier transform of the product $\mathcal{F}_x\big[G^1_0\big]\,\mathcal{F}_x\big[S_{\xi}\big]$. 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 zero 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 $1/(2\xi)$, and so

$$V_\xi(u) = \left(u \bmod \frac{1}{2\xi}\right) - \frac{1}{4\xi} \tag{54}$$

is a suitable distance function for the cosine signal (51). The case of sine signals is analogous, with sin replaced by cos in the linear response (3.2.3), and u replaced by $u - 1/(4\xi)$ in the distance function (54).

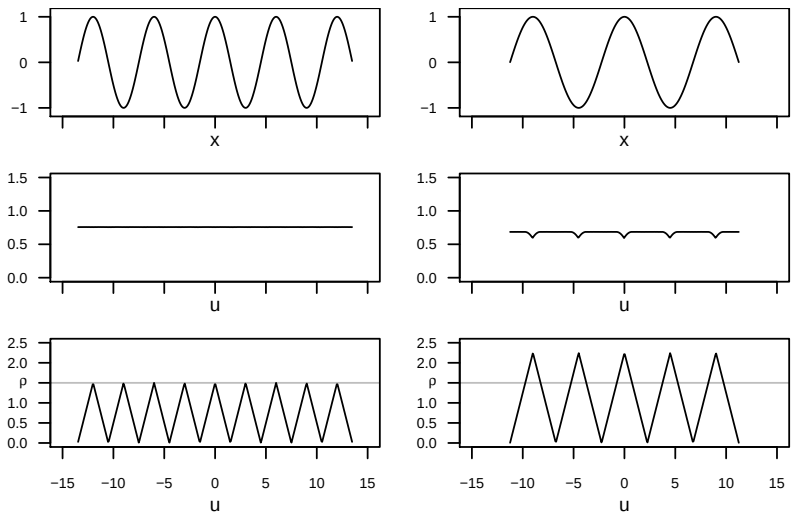

Figure 7: Cosine response. **Left column**: The top plot shows the cosine signal $S_\xi(x)$, as in (51), of frequency $\xi = 1/6$. The middle plot shows the response $C(u)$, which is constant. The bottom plot shows the distance function $|V_\xi(u)|$, as in (54). Note that the critical case is plotted, 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 before, but for a lower frequency, $\xi = 1/9$. 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 (54), with

respect to u, are $\pm 1/(4\xi)$, from which it follows that the response is identically $R^\star_\xi$ if $\xi \geq 1/(4\rho)$. The corresponding constraint on the wavelength $1/\xi$ is

$$1/\xi \leq 4\rho. \tag{55}$$

In order to interpret this result, recall that $\rho \geq \sigma$ is required for a unimodal impulse response (42). 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 $\xi = 1/(6\sigma)$. The response level is $R^\star_\xi = 0.759$ at the critical frequency, as can be confirmed in figure 7.

3.2.4 Gaussian

The fourth signal to be considered is the Gaussian pulse, which models a symmetric blob in the image. The pulse is defined in the un-normalized form

$$S_\tau(x) = G(x, \tau) \tag{56}$$

so that the maximum value is $S_\tau(0) = 1$. The convolution of $G^1_1(x, \sigma)$ and $S_\tau(x)$ results in another Gaussian, of variance $\sigma^2 + \tau^2$. The linear response can be expressed as

$$R_\tau(0, u) = \pi\sigma\tau\, G^0_1\left(u, \sqrt{\sigma^2 + \tau^2}\right). \tag{57}$$

The extrema of this function can be found by differentiating $R_\tau(0, u)$, and setting the result to zero. The derivative contains a factor $\sigma^2 + \tau^2 - u^2$, and so there are solutions $u = \pm\sqrt{\sigma^2 + \tau^2}$. The maximum corresponds to the negative argument, hence

$$R^\star_\tau = \pi\sigma\tau\, G^0_1\left(-\sqrt{\sigma^2 + \tau^2}, \sqrt{\sigma^2 + \tau^2}\right). \tag{58}$$

The two extrema, derived above, correspond to the maxima of $|R_\tau(0, u)|$. The distance function can therefore be defined as

$$V_\tau(u) = u - \mathrm{sgn}_+(u)\sqrt{\sigma^2 + \tau^2}. \tag{59}$$

If $\rho = \sqrt{\sigma^2 + \tau^2}$, then the plateaus associated with the left and right extrema meet in the middle, meaning that there is a single interval over which $R(u) = R^\star_\tau$. It has been established that, in practice, $\rho > \sigma$. It follows that the unimodal range is

$$\tau \leq \sqrt{\rho^2 - \sigma^2}. \tag{60}$$

If the standard settings $\sigma = 1$ and $\rho = 1.5\sigma$ are used, then it follows from (58) and (60) that $\tau = 1.118$ and $R^\star_\tau = 0.378$ in the critical case.

It may be noted that the Gaussian impulse S_τ can be used to represent the Dirac δ-distribution in the limit $\tau \to 0$. The above definitions continue to apply, with τ set to zero; indeed, the results in section 3.2.1 are special cases of the above analysis.

3.2.5 Bar

The fifth and final signal to be considered is the *bar*, of width 2λ. This can be represented by the difference of two step-functions,

$$S_\lambda(x) = \tfrac{1}{2}\big(\mathrm{sgn}(x + \lambda) - \mathrm{sgn}(x - \lambda)\big) \tag{61}$$

It follows, with reference to (49) that the initial response is the sum of two Gaussians, separated by 2λ:

$$R_\lambda(u,0) = \tfrac{1}{2}\big(G(u+\lambda,\sigma) + G(u-\lambda,\sigma)\big) \tag{62}$$

It is difficult to obtain analytic results in this case, owing to the form of (62). However, note that for moderate values of λ, the bar response can be treated as the combination of two independent step responses. In this case the maxima would be at $\pm\lambda$ and so, setting $u = \lambda$, it follows that

$$R^\star_\lambda \approx \tfrac{1}{2}\big(G(0,\sigma) + G(2\lambda,\sigma)\big). \tag{63}$$

The distance function can now be defined by analogy with (59), again assuming that the peaks are at $\pm\lambda$, hence

$$V_\lambda(u) \approx u - \mathrm{sgn}_+(u)\,\lambda. \tag{64}$$

If $\lambda \approx \rho$ then the plateaus associated with the two peaks will meet in the middle. This situation is analogous to that of the Gaussian pulse, discussed above. It follows that the unimodal response range is approximately

$$\lambda \leq \rho. \tag{65}$$

The accuracy of these approximations can be confirmed numerically. For example, consider the critical case $\lambda = \rho$, corresponding to the widest bar that could give a unimodal response. In this case the true maxima of (62) are the roots of $G_1(u+\rho,\sigma) + G_1(u-\rho,\sigma)$, which can be found by standard numerical methods (Press et al., 1992). The standard settings $\sigma = 1$ and $\rho = 1.5\sigma$ were used, and the true extrema were found at $u = \pm 1.463$. This corresponds to a relative error of 2.45%, with respect to the assumed locations $\pm\lambda$. Finally, the response constant is $R^\star_\lambda = 0.506$ in this case.

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), as described in section 1.3. The procedure is as follows. Each 1024×768 greyscale 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 $5°$. 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

$$\boldsymbol{x}_{ijk} = \boldsymbol{p}_j \pm k\Delta \times (\cos\theta_i, \sin\theta_i)^\top, \quad \text{where} \quad j = 1, \ldots, m \quad \text{and} \quad k = 0, \ldots, n \tag{66}$$

is sampled from each 2-D 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 one 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 non-integral positions $\boldsymbol{x}_{ijk}$ are obtained by bilinear interpolation. The samples are non-negative 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,

$$E_S(i) = \frac{1}{mn} \sum_{j=1}^{m} \sum_{k=0}^{n} \gamma S_1\big(\boldsymbol{x}_{ijk}, \sigma, \theta_i\big) \tag{67}$$

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;

$$Q_S(i) = \frac{1}{mn} \sum_{j=1}^{m} \sum_{k=1}^{n} \Big|\gamma S_1\big(\boldsymbol{x}_{ijk}, \sigma, \theta_i\big) - \gamma S_1\big(\boldsymbol{x}_{ij[k-1]}, \sigma, \theta_i\big)\Big|^2. \tag{68}$$

The coordinates $\boldsymbol{x}_{ijk}$ and $\boldsymbol{x}_{ij[k-1]}$ represent adjacent points (separated by Δ) on the j-th path in the i-th 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, as described in section 1.3. 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. Firstly, a vertical cosine grating, $I_{\cos}(x, y) = \frac{1}{2}\big(1 + \cos(2\pi\xi x)\big)$. The range is set to $\rho = 1.5\sigma$, as usual, and the wavelength is set to $1/\xi = 8\sigma$. These values do *not* satisfy the limit (55), which ensures that the complex response will not be trivial. The simple-cell response is shown in figure 8 (top left), and two effects should be noted. Firstly, the response is tuned to the dominant orientation $\theta = 0$, as can be seen from the unimodal shape of the curve. Secondly, 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). It can be seen that 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 2-D. The complex response reduces the variation, as shown in figure 8 (bottom).

Figure 9 (top) shows results for a real image (fig. 10, top) which has an orientation-structure similar to that of the grating. The results are analogous. The 2-D simple and complex responses, in this case, are visualized in figure 10. It can be seen that the complex response shows much less variation in the direction orthogonal to the image structure. 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.

Finally, it is interesting to consider the statistics of the simple and complex responses in this model. The simple response, which is the (absolute) Gaussian derivative of the image, is already well-studied. In particular, it is found that log-histogram has a characteristic apex-shape, indicating high kurtosis (Simoncelli

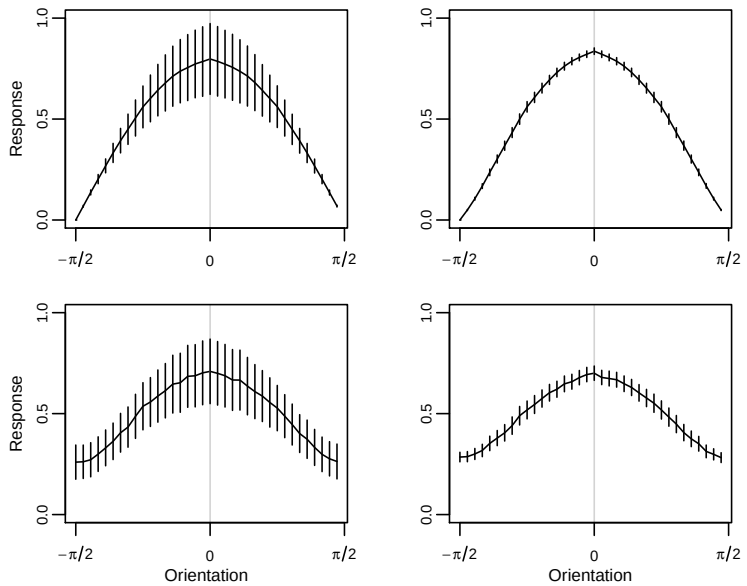

Figure 8: Cosine response. **Top left:** Average simple cell response $E_S(i)$ to a vertical cosine grating of wavelength 8σ. The curve indicates the mean response in each of 36 orientation channels θ_i, and has a clear peak at zero. The vertical bars $\pm\frac{1}{2}\sqrt{Q_S(i)}$ indicate the RMS spatial 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 Olshausen, 2001). The complex response, in the differential model, splits the histogram at the peak, because there is *some* directional variation within $\pm\rho$ of almost every point in a natural image. This effect is shown in figure 11.

It is also interesting to consider the autocorrelation of the responses. As shown in figures 8 and 9, the complex response varies less rapidly than the simple response. This can also be seen in the directional autocorrelation functions shown in figure 12. It is particularly interesting to look at this in 2-D, for a natural image which contains no dominant orientation. The simple response $S_1(\boldsymbol{x}, \sigma, \theta)$ increases the variability in the direction θ, as before. Figure 12, however, shows that the complex response tends to be isotropic within each orientation band.

There is an interesting relationship between the response histograms and autocorrelation responses. The histograms show that the mean response (and the number of non-zero responses) is higher in the complex representation. However, the autocorrelation functions show that the complex response varies more slowly, and more isotropically. This suggests that it could be *subsampled*, while preserving information about the local structure of the image.

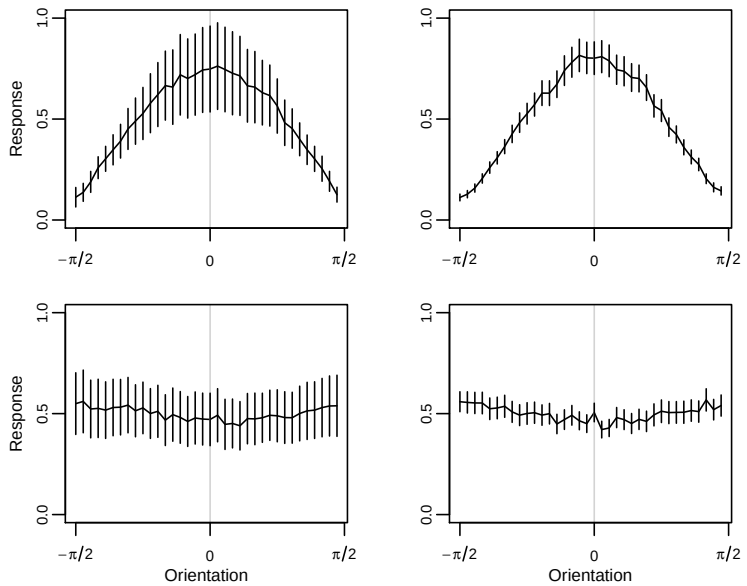

Figure 9: Image response. **Top:** As in figure 8, but using a real image that contains a dominant vertical orientation (fig. 10, top). Left: The simple response shows variation across all orientation channels $Q_S(i)$. Right: The variation of the complex response $Q_C(i)$ is much lower. **Bottom:** As before, but using a natural image, with no dominant orientation.

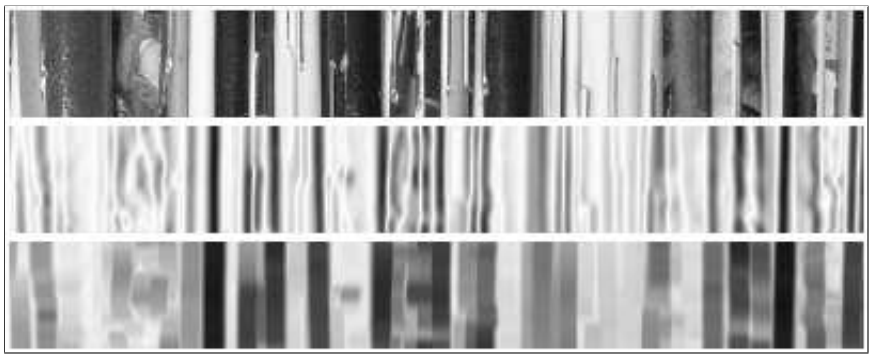

Figure 10: A 2-D example. **Top:** A slice through a real image (of a stack of magazines) that has a dominant orientation-structure. **Middle:** A slice through the simple cell response $S_1(\boldsymbol{x}, \sigma, \theta_i)$ in the dominant orientation channel. Dark intensity indicates a strong response (due to an edge-like structure). The response varies greatly in the horizontal direction, which corresponds to the long bars around the peak in figure 9 (top left). **Bottom:** A slice through the complex response $C(\boldsymbol{x}, \sigma, \theta_i)$, which shows much less variation, as in figure 9 (top right).

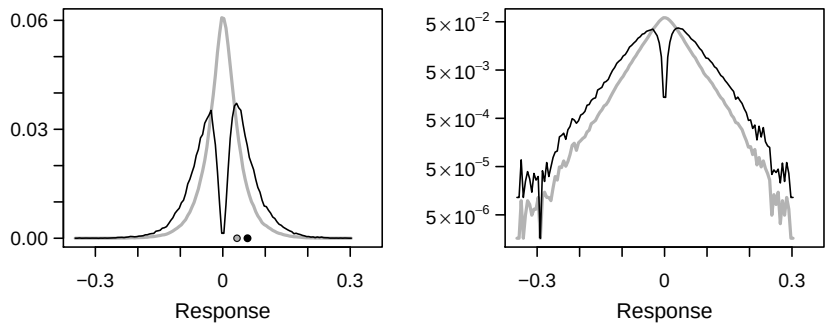

Figure 11: Response histograms. **Left:** The grey line shows the histogram of first-derivative filter responses $S_1(\boldsymbol{x}, \sigma, \theta)$, with $\sigma = 2$ and $\theta = 0$. The black line shows the corresponding complex response $C(\boldsymbol{x}, \sigma, \theta)$, using $\rho = 1.5\sigma$. The two dots indicate the mean absolute values of the corresponding response. **Right:** The same histograms, plotted on a logarithmic scale. Note the characteristic apex-shape of the S_1 curve, in grey.

4 Discussion

It has been shown that a differential model of the complex cell can be constructed from the local jet representation. This can be viewed as a formalization of the Hubel and Wiesel (1962) subunit model . It has been shown that the new model works naturally with a basis of steerable filters in 2-D. The qualitative components in the present approach are similar to those of the Gabor energy model. Both models are based on oriented linear filters, which are centred at the same position. Likewise, both models output a combination of the nonlinearly transformed filter responses.

It will be important, in future work, to make a *quantitative* comparison between the present model and the existing Gabor energy approach. This task is not straightforward, and is beyond the scope of the present work. The relationship between the two models is complicated, owing to their different parameterizations and abilities. The Gabor model is parameterized by continuous scale σ, orientation θ, and frequency ξ. It is most easily analyzed in relation to sinusoidal signals at the given frequency. The differential model is parameterized by scale σ, orientation θ, and *integer* order, k. It is most easily analyzed in relation to step-edges, which contain all frequencies. Hence, for sinusoidal signals at frequency ξ, the Gabor model requires just two filters (in quadrature) and is therefore much more compact. For more general signals, however, the Gabor mechanism must be replicated across frequencies, and so the number of filters in each model becomes comparable. Furthermore, there is no generally agreed way to integrate across frequencies in the energy model. It would be necessary to resolve these issues before making a quantitative comparison. This task is left for future work.

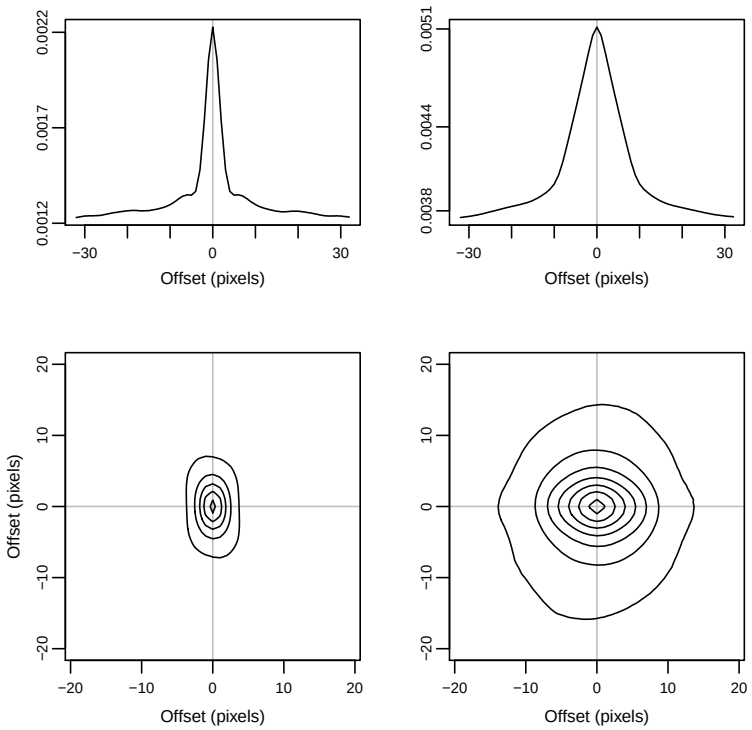

Figure 12: Response autocorrelation. **Top left:** One-dimensional autocorrelation of the response $S_1(\boldsymbol{x}, \sigma, \theta)$ in direction $\theta = 0$. **Top right:** Corresponding autocorrelation of the complex response $C(\boldsymbol{x}, \sigma, \theta)$. Note the greater width of the second function. **Bottom left:** Two-dimensional autocorrelation of the response $S_1(\boldsymbol{x}, \sigma, \theta)$, with $\theta = 0$. Note the anisotropy, which indicates that the response changes more rapidly in the direction of the derivative. **Bottom right:** Corresponding autocorrelation (with the same contour-separation) of the complex response $C(\boldsymbol{x}, \sigma, \theta)$. Note the greater isotropy of the second function.

4.1 Predictions

The differential model makes several predictions about the configuration of simple and complex cells. Firstly, 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. Secondly, the differential model 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, as discussed in the following section, the complex receptive fields can be extended by allowing multiple scales σ_j in the basis set of the differential model.

Perhaps the most interesting prediction of the present model is that high-order derivative filters are required, in order to approximate the target filter over a sufficient range ρ. In particular, it was shown 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 five and beyond must be used in the approximation, as can be seen from figure 4. This is interesting, because very oscillatory filters have been observed in V1 (Young and 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. The present work suggests that these filters could have a different role, in providing a basis for spatially offset filters of low order. Lastly, the isotropy of the complex autocorrelation functions, as in figure 12 is interesting. It appears that the same isotropic spatial sampling could be used across all orientation channels, without over-representing any direction. This would be consistent with the local-jet model, in which all orientations are represented at each spatial position.

4.2 Extensions

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 (with odd, even & mixed symmetry), 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 et al., 1987; Sasaki and Ohzawa, 2007; Touryan et al., 2005). Neither of these extensions requires any modification of the present theory or algorithms. It would also be interesting to consider the relationship of the present work to existing models of motion and binocular disparity processing (Fleet et al., 1996; Johnston et al., 1992).

Another direction would be to consider how the differential model could be *learned* from natural image data, by analogy with the models described in section 1.3. 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, by analogy with slow-feature analysis models (Berkes and Wiskott, 2005; Wiskott and Sejnowski, 2002). The combination of the geometric and statistical approaches to image analysis is, more generally, a very promising aim.

Appendices

A The Local Jet Representation

This section summarizes the spatial image-representation, formulated by Koenderink and van Doorn (1987), which motivates the use of Gaussian derivative

filters. If $S(x, y)$ denotes the luminance of the original image, then the blurred luminance $S(x, y, \sigma) = G(x, y, \sigma) \star S(x, y)$ can be expanded in a Taylor series, with respect to the Cartesian coordinates (x, y), as follows:

$$S(x+u, y+v, \sigma) = \sum_{k}^{\infty} \frac{1}{k!} \left(u\frac{\partial}{\partial x} + v\frac{\partial}{\partial x} \right)^k S(x, y, \sigma) \tag{69}$$

$$= \sum_{k}^{\infty} \sum_{\ell}^{k} \binom{k}{\ell} \frac{u^{k-\ell} v^{\ell}}{k!} S_{k\ell}(x, y, \sigma) \tag{70}$$

The coefficients $S_{k\ell}(x, y, \sigma)$ are the derivatives of the blurred luminance, in Cartesian coordinates, which can be obtained by convolution with the filters

$$S_{k\ell}(x, y, \sigma) = \frac{\partial^{k+\ell}}{\partial x^k \partial y^\ell} S(x, y, \sigma) \tag{71}$$

$$= G_{k\ell}(x, y, \sigma) \star S(x, y). \tag{72}$$

If the series (70) is truncated at some order K, then the *local jet* is the set of images with the *same* coefficients $S_{k\ell}$, up to order K inclusive. The corresponding collection of partial derivatives encodes the intrinsic geometry of the jet. Furthermore, the filters $S_{k\ell}$ can be re-combined into operators that compute specific geometric properties (e.g. the curvature of an edge).

This paper is largely concerned with the *directional* structure of the image, which means that the offsets (u, v) in (70) can be replaced by a single variable λ, where $(u, v) = \lambda(\cos\theta, \sin\theta)$ for a given θ. It then suffices to consider the 1-D expansion with coefficients obtained from $G_k(x, y, \sigma, \theta) \star S(x, y)$, owing to the separability of the Gaussian derivatives.

B The Gabor Energy Model

This section summarizes the Gabor energy model, based on the account given by Dayan and Abbot (Dayan and Abbott, 2001). Let F and $F_\perp$ be Gabor filters that differ in phase by $\pi/2$,

$$\begin{aligned} F(x, \xi, \sigma) &= G_0(x, \sigma)\, \cos(2\pi\xi x) \\ F_\perp(x, \xi, \sigma) &= G_0(x, \sigma)\, \sin(2\pi\xi x) \end{aligned} \tag{73}$$

where ξ is the frequency and σ is the standard deviation of the Gaussian envelope. Consider a sinusoidal signal of phase ϕ and frequency η,

$$S(x, \eta, \phi) = \alpha \cos\bigl(2\pi\eta x + \phi\bigr). \tag{74}$$

The response of each filter to the signal component is obtained from the inner product with S. For example, the response of the even filter is

$$F \cdot L = \int_{-\infty}^{\infty} F(x, \xi, \sigma)\, S(x, \eta, \phi)\, \mathrm{d}x \tag{75}$$

which is also the value of the convolution $F \star S$ at the origin. The integrals $F \cdot S$ and $F_\perp \cdot S$ are as follows:

$$\begin{aligned} F \cdot S &= \frac{\alpha}{2} \cos(\phi) \Big(\epsilon + \exp\big(-2\pi^2\sigma^2(\xi-\eta)^2\big) \Big) \\ F_\perp \cdot S &= \frac{\alpha}{2} \sin(\phi) \Big(\epsilon - \exp\big(-2\pi^2\sigma^2(\xi-\eta)^2\big) \Big) \\ &\text{where } \epsilon = \exp\big(-2\pi^2\sigma^2(\xi+\eta)^2\big) \approx 0. \end{aligned} \tag{76}$$

The assertion $\epsilon \approx 0$ is justified by expanding the quadratics $(\xi \pm \eta)^2$, and noting that $\exp\big(-\zeta^2\big) \pm \exp\big(\zeta^2\big) \approx \exp\big(\zeta^2\big)$ where ζ is the cross-term $4\pi^2\sigma^2\xi\eta$. This approximation is valid if the spatial frequencies of the signal and filter, with respect to σ, are not too low.

It can be seen, after setting $\epsilon = 0$, that the filter responses (76) are products of three terms; the half-amplitude, a cosinusoidal or sinusoidal function of phase, and a Gaussian function of the frequency difference. If $\phi = 0$ then $F \cdot S$ is maximal, while $F_\perp \cdot S$ is zero. Conversely, if $\phi = -\pi/2$, then $F \cdot S$ is zero, while $F_\perp \cdot S$ is maximal.

The final energy response is obtained by squaring and summing the responses $F \cdot S$ and $F_\perp \cdot S$. This leads, via the trigonometric identity $\cos^2(\phi) + \sin^2(\phi) = 1$ to the output

$$\big(F \cdot S\big)^2 + \big(F_\perp \cdot S\big)^2 \approx \frac{\alpha^2/4}{\exp\big(4\pi^2\sigma^2(\xi-\eta)^2\big)} \tag{77}$$

which is independent of the signal phase ϕ. It should be emphasized, however, that a periodic response is obtained for stimulus-components that do not match the filter frequency ξ.

References

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

Alonso, J.-M. and 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. and 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. and 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 Fougelman-Soulie, F. and Hérault, J., editors, *Neuro-computing: Algorithms, Architectures and Applictions*. 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 and complex cells compute. *J. Physiology*, 577(2):463–466.

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

Carandini, M., Hegger, D. J., and Movshon, J. A. (1997). Linearity and Normalization in Simple Cells of the Macaque Primary Visual Cortex. *J. Neuroscience*, 17(21):8621–8644.

Chance, F. S., Nelson, S. B., and Abbott, L. F. (1999). Complex cells as cortically amplified simple cells. *Nature Neuroscience*, 2(3):277–282.

Cumming, B. G. and DeAngelis, G. C. (2001). The physiology of stereopsis. *Ann. Rev. of Neurosci.*, 24:203–238.

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. and Abbott, L. F. (2001). *Theoretical Neuroscience*. MIT Press.

Dobbins, A., Zucker, S. W., and 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., and 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. and Sommer, G. (2001). The Monogenic Signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144.

Ferster, D. and Miller, K. D. (2000). Neural Mechanisms of Orientation Selectivity in the Visual Cortex. *Annual Review of Neuroscience*, 23:441–471.

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

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, 3:194–200.

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

Gaska, J. P., Pollen, D. A., and 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., and Hesse, G. S. (2007). From filters to features: Scale-space analysis of edge and blur coding in human vision. *Journal of Vision*, 7(13):1–21.

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

Hawken, M. J. and 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. (1988). Optic flow using spatiotemporal filters. *International Journal of Computer Vision*, 1(4):279–302.

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

Hochstein, S. and Shapley, R. M. (1976). Linear and nonlinear spatial subunits in Y catretinal ganglion cells. *J. Physiology*, 262:265–284.

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

Hyvärinen, A. and Hoyer, P. O. (2000). Emergence of Phase- and Shift-Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces. *Neural Computation*, 12:1705–1720.

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

Johnston, A., McOwan, P. W., and 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. and 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. and Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, 457:83–86.

Kjaer, T. W., Gawne, T. J., Hertz, J. A., and 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. and van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375.

Körding, K. P., Kayser, C., Einhäuser, W., , and König, P. (2004). How Are Complex Cell Properties Adapted to the Statistics of Natural Stimuli? *J. Neurophysiology*, 91:206–212.

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

Martinez, L. M. and Alonso, J.-M. (2003). Complex Receptive Fields in Primary Visual Cortex. *The Neuroscientist*, 9(5):317–331.

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

Mechler, F., Reich, D. S., and Victor, J. D. (2002). Detection and Discrimination of Relative Spatial Phase by V1 Neurons. *J. Neuroscience*, 22(14):6129–6157.

Mechler, F. and Ringach, D. L. (2002). On the Classification of Simple and Complex Cells. *Vision Research*, 42:1017–1013.

Mel, B. W., Ruderman, D. L., and Archie, K. A. (1998). Translation-Invariant Orientation Tuning in Visual Complex Cells Could Derive from Intradendritic Computations. *J. Neuroscience*, 18(11):4325–4334.

Movshon, J. A., Thompson, I. D., and 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., and 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.

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

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C.* Cambridge University Press, 2nd edition.

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

Sanger, T. D. (1988). Stereo disparity computation using Gabor filters. *Biological Cybernetics*, 59(6):405–418.

Sasaki, K. S. and Ohzawa, I. (2007). Internal Spatial Organization of Receptive Fields of Complex Cells in the Early Visual Cortex. *J. Neurophysiology*, 98:1194–1212.

Serre, T., Oliva, A., and Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc. Nat. Acad. Sci. USA*, 104(15):6424–6429.

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

Simoncelli, E. P. and Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research*, 38(5):743–761.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216.

Skottun, B. C., Valois, R. L. D., Grosof, D. H., Movshon, J. A., Albrecht, D. G., and Bonds, A. B. (1991). Classifying Simple and Complex Cells on the basis of Response Modulation. *Vision Research*, 31(7/8):1079–1086.

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

Tao, L., Shelley, M., McLaughlin, D., and Shapley, R. (2004). An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proc. Nat. Acad. Sci. USA*, 101(1):366–371.

Touryan, J., Felsen, G., and Dan, Y. (2005). Spatial Structure of Complex Cell Receptive Fields Measured with Natural Images. *Neuron*, 45:781–791.

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

Wundrich, I. J., von der Malsburg, C., and Würtz, R. P. (2004). Image Representation by Complex Cell Responses. *Neural Computation*, 16(4):2563–2575.

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

Yuille, A. L. and Geiger, D. (2003). Winner-Take-All Networks. In Arbib, M. A., editor, *The Handbook of Brain Theory and Neural Networks*. MIT Press.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399