A multi-layer sparse coding network learns contour coding from natural images*

Patrik O. Hoyer and Aapo Hyvärinen
Neural Networks Research Centre, Helsinki University of Technology
P.O. Box 9800, FIN-02015 HUT, Finland
patrik.hoyer@hut.fi

To appear in *Vision Research*
January 3, 2002

Abstract

An important approach in visual neuroscience considers how the function of the early visual system relates to the statistics of its natural input. Previous studies have shown how many basic properties of the primary visual cortex, such as the receptive fields of simple and complex cells and the spatial organization (topography) of the cells, can be understood as efficient coding of natural images. Here we extend the framework by considering how the responses of complex cells could be sparsely represented by a higher-order neural layer. This leads to contour coding and end-stopped receptive fields. In addition, contour integration could be interpreted as top-down inference in the presented model.

Keywords: natural images, neural networks, contours, cortex, independent component analysis

1 Introduction

1.1 Why build statistical models of natural images?

After Hubel and Wiesel (1962; 1968) first showed that neurons in mammalian primary visual cortex (V1) are optimally stimulated by bars and edges, a large part of visual neuroscience has been concerned with exploring the response characteristics of neurons in V1 and in higher visual areas. However, such studies do not directly answer the question of why the neurons respond in the way that they do. Why does it make sense to filter the incoming visual signals with receptive fields such as those of V1 simple cells? What is the goal of the neural code? Not only are such questions interesting in their own right, but finding answers would give us a deeper understanding of the information processing in the visual system and could even give predictions for neuronal receptive fields in higher visual areas.

One important approach for answering such questions is to consider how the function of the visual system relates to the properties of natural images. It has long been hypothesized that the early visual system is adapted to the input statistics (Attneave, 1954; Barlow, 1961). Such an adaptation is thought to be the result of the combined forces of evolution and neural learning during development. This hypothesis has lately been gaining ground as information-theoretically efficient coding strategies have been used to explain much of the early processing of visual sensory data, including response properties of neurons in the retina (Srinivasan et al., 1982; Atick and Redlich, 1992), lateral geniculate nucleus (Dong and Atick, 1995; Dan et al., 1996), and V1 (Olshausen and Field, 1996; Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998; van Hateren and Ruderman, 1998; Rao and Ballard, 1999; Simoncelli and Schwartz, 1999;

*This paper previously had the title: 'A non-negative sparse coding network learns contour coding and integration from natural images'. The title was changed during revision.

Zetzsche and Krieger, 1999; Hoyer and Hyvärinen, 2000; Hyvärinen and Hoyer, 2000; Tailor et al., 2000; Hyvärinen and Hoyer, 2001; Wachtler et al., 2001). For a recent review, see (Simoncelli and Olshausen, 2001).

Although there seems to be a consensus that information-theoretic arguments are relevant when investigating the earliest parts of the visual system, there is no general agreement on how far such arguments can be taken. Can information theory be used to understand neuronal processing higher in the processing 'hierarchy', in say, areas V2, V4, or perhaps even inferotemporal cortex? One might, for instance, be inclined to think that higher processes in the visual system would be expected to be concerned with specific tasks (such as the estimation of shape or heading) and not simply 'representing the information efficiently'. Early levels, the argument goes, cannot be very goal-oriented (as there are a number of tasks that need to use that information) and thus must simply concentrate on representing the information faithfully and efficiently.

We would like to suggest that emphasis should be placed not on the hypothesis that the cortex simply seeks to represent the sensory data efficiently, but rather on the notion that it builds a probabilistic internal model for that data. Such a change in emphasis from the original redundancy reduction hypothesis has recently been compellingly argued for by Barlow (2001a; 2001b).[a]

In such a framework, it is natural to think of neural networks not as simply transforming the input signals into coded representations having some desired properties (such as sparseness), but rather as *modeling* the structure of the sensory data. Viewed in this light, data-driven probabilistic models make as much sense at higher levels of the processing hierarchy as at the earliest stages of the visual system. Although not being the focus of the large majority of work on sensory coding (reviewed in Simoncelli and Olshausen, 2001), this point of view has nonetheless been emphasised by a number of researchers (see e.g. (Mumford, 1994; Dayan et al., 1995; Hinton et al., 1995; Hinton and Ghahramani, 1997; Olshausen and Field, 1997; Rao and Ballard, 1999)) and is also the approach we take.

1.2 Modeling V1 receptive fields with sparse coding

In an influential paper, Olshausen and Field (1996) showed how the classical receptive fields of simple cells in V1 can be understood in the framework of sparse coding. The basic idea is to model the observed data (random variables) x_j as a weighted sum of some hidden (latent) random variables s_i, to which Gaussian noise has been added:[b]

$$x_j = \sum_{i=1}^{n} a_{ji} s_i + n_j. \tag{1}$$

This can be expressed compactly in vector form (with bold letters indicating vector quantities) as

$$\mathbf{x} = \sum_{i=1}^{n} \mathbf{a}_i s_i + \mathbf{n}. \tag{2}$$

In other words, each observed data pattern $\mathbf{x}$ is approximately expressed as some linear combination of the basis patterns $\mathbf{a}_i$. The hidden variables s_i that give the mixing proportions are stochastic and differ for each observed $\mathbf{x}$.

The crucial assumption in the sparse coding framework is that the hidden random variables s_i are mutually independent and that they exhibit sparseness. Sparseness is a property independent of scale (variance), and implies that the s_i have probability densities which are highly peaked at zero and have heavy tails. Essentially, the idea is that any single typical input pattern $\mathbf{x}$ can be accurately described using only a few active (significantly non-zero) units s_i. However, all of the basis patterns $\mathbf{a}_i$ are needed to represent the data because the set of active units changes from input to input.

[a]Of course, information theory tells us that the better model we have for some given data, the more efficiently (compactly) we could potentially represent it. However, for the case of the brain, it may well be that it is the model that is important, not the forming of a compact representation.

[b]This sparse coding model is also called the noisy Independent Component Analysis (ICA) model (Hyvärinen et al., 2001).

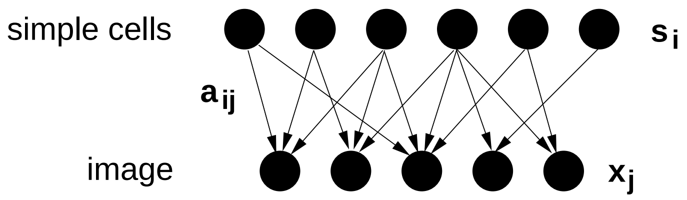

Figure 1: The linear sparse coding neural network. Units are depicted by filled circles, and arrows represent conditional dependencies (in the generative model) between the units. Upon observing data $\mathbf{x}$ the hidden neuron activities s_i are calculated as the most probable latent variable values to have generated the data. On a longer timescale, the generative weights a_{ij} are adapted to allow typical data to be represented sparsely.

This model can be represented by a simple neural network (see Figure 1), where the observed data $\mathbf{x}$ (represented by the activities in the lower, input layer) is a linear function of the activities of the hidden variables s_i in the higher layer, contaminated by additive Gaussian noise (Olshausen and Field, 1997). Upon observing an input $\mathbf{x}$, the network calculates the optimal representation s_i in the sense that it is the configuration of the hidden variables most likely to have caused the observed data. This is *inferring* the latent variables, and is the short timescale goal of the network. Since the prior probability of the s_i is sparse, this boils down to finding a maximally sparse configuration of the s_i that nevertheless approximately generates $\mathbf{x}$. In the long run, the goal of the network is to *learn* (adapt) the generative weights (basis patterns) $\mathbf{a}_i$ so that the probability of the data is maximized. Again, for sparse latent variables, this is achieved when the weights are such that only a few higher layer neurons s_i need to be significantly active to represent typical input patterns.

When the network is trained on data consisting of patches from natural images, with the elements of $\mathbf{x}$ representing pixel gray-scale values, the learned basis patterns $\mathbf{a}_i$ resemble Gabor functions and V1 simple cell classical receptive fields, and the corresponding s_i can be interpreted as the activations of the cells (Olshausen and Field, 1996; Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998). This basic result has already been extended to explain spatiotemporal (van Hateren and Ruderman, 1998), chromatic (Hoyer and Hyvärinen, 2000; Tailor et al., 2000; Wachtler et al., 2001), and binocular (Hoyer and Hyvärinen, 2000) properties of simple cells. This can be done simply by training the network on input data consisting of image sequences, colour images, and stereo images, respectively.

Although the basic sparse coding model has been quite successful at explaining the receptive fields of simple cells in V1, it is not difficult to see that it cannot account for the behaviour of V1 complex cells. These cells are, just as simple cells, sensitive to the orientation and spatial frequency of the stimulus, but unlike simple cells they are not very sensitive to the phase of the stimulus. Such *invariance* is impossible to describe with a strictly linear model. Consider, for instance, reversing of contrast polarity of the stimulus. Such an operation would flip the sign of a linear representation, whereas the response of a typical complex cell does not change to any significant degree.

To model such stimulus invariance, Hyvärinen and Hoyer (2000) modified the network to include higher-order cells that pooled the energies (squared outputs) of groups of simple cells. Such pooling can be interpreted as a generative model in which the higher-order cells determine the variances of the simple cells (Hyvärinen et al., 2001). This model is depicted in Figure 2. (Note that simply adding a second linear layer on top of the first one would not be very useful, because a linear transform followed by another linear transform is still a linear transform, so nothing is gained by a multi-layer network.) When trained on natural image patches, the adaptable weights $\mathbf{a}_i$ converged to simple cell-like receptive fields, and the behaviour of the higher-order units was qualitatively similar to complex cell responses.[c] This behaviour was due to the fact that simple cells in any single group learned receptive fields of similar orientation and

[c]The model actually estimated in (Hyvärinen and Hoyer, 2000) was a simplified version where there was no noise and the number of hidden variables s_i was equal to the dimensionality of the data.

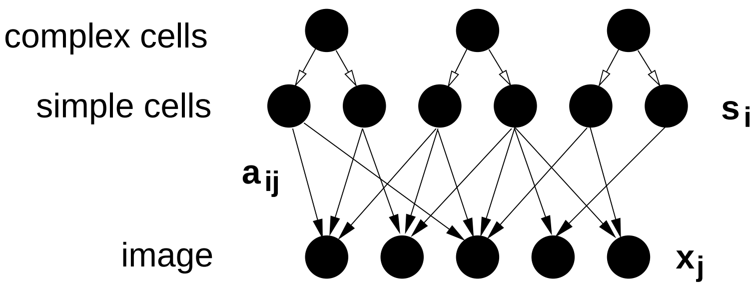

Figure 2: The extended sparse coding model (Hyvärinen and Hoyer, 2000; Hyvärinen et al., 2001). Each complex cell determines the *variance* of a group of simple cells.

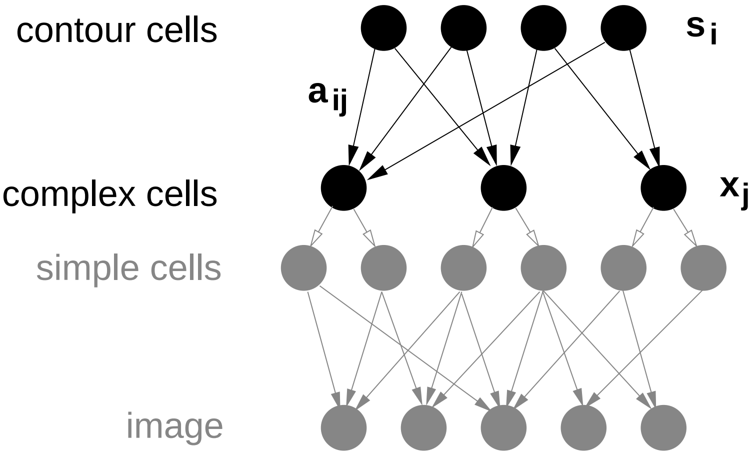

Figure 3: The simplified hierarchical model investigated in this paper. Model complex cell responses are calculated in a feedforward manner, and these responses are subsequently analyzed by a higher-order sparse coding layer in the network. To emphasise that the lower layers are fixed and not learned, these layers have been grayed out in the figure.

spatial frequency, but differing in spatial phase. A further extension was introduced in (Hyvärinen and Hoyer, 2001; Hyvärinen et al., 2001) where it was shown how V1-like topography can additionally emerge.

1.3 A multi-layer model

An important question is how to extend these models to account for response properties of neurons higher in the processing hierarchy. Perhaps the most straightforward approach is to add a linear layer on top of the complex cell model. This would amount to assuming a model where the activities of our model complex cells are not independent, but rather can be described as a linear combination of some higher-order independent units. In this contribution we study a simplified version of that model, where the lower layers are fixed and the responses of the model complex cells are a straightforward function of the image input. This situation is depicted in Figure 3, where the lower layers have been grayed out to illustrate that these layers are not an active (learned) component in this simplified model.

The choice to investigate the simplified model was driven by several factors. First, this model is computationally significantly simpler to learn, as only one set of weights needs to be adapted. Thus, experiments can be performed at a larger scale. Also, the fact that the operation of the model complex cells is completely specified makes the interpretation of the results more straightforward than in the full

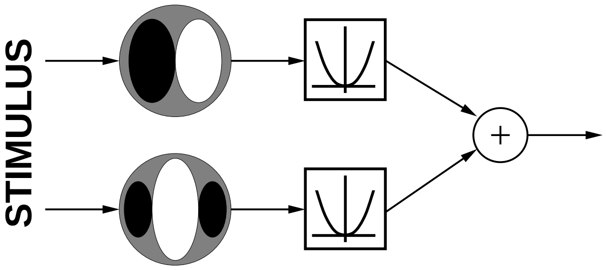

Figure 4: Complex cell model used. The response of a complex cell was obtained by linearly filtering with quadrature gabor filters, taking squares, and summing.

model: Fixing the lower layer structure allows for a simple visualization and analysis of the results, compared with an unrestricted model. Finally, using the chosen feedforward complex cell response model allows our results to be compared to other recent work (Krüger, 1998; Sigman et al., 2001; Geisler et al., 2001) analyzing the dependencies of complex cell responses. We believe that the analysis provided in this paper can be viewed as a preliminary investigation into how complex cell responses could be represented in an unrestricted multi-layer model.

In brief, we model V1 complex cell outputs by a classic complex cell energy model and, using these responses as the input $\mathbf{x}$, estimate the linear model of Eq. (2) assuming sparse, non-negative s_i. We show how our network learns contour coding from natural images in an unsupervised fashion, and discuss how contour integration could be viewed as resulting from top-down inference in the model.

2 Methods

2.1 Model complex cells

We modeled complex cell responses by a very simple and widely used energy model (Pollen and Ronner, 1983; Adelson and Bergen, 1985; Morrone and Burr, 1988), detecting spatially localized oriented Fourier energy in static, monocular, achromatic images. The response of a model complex cell was calculated by summing the squared responses of a pair of quadrature gabor filters. (For details, see the Appendix.) This is depicted in Figure 4.

For simplicity of interpretation and for computational reasons, we restricted our analysis to a single spatial scale, and the cells were placed on a rectangular 6-by-6 grid with 4 differently oriented cells at each location. In Figure 5 we illustrate the behaviour of our model complex cells by showing their responses to a few natural image patches. The figure also shows the response distribution (over the ensemble of image patches) of a single cell. The distribution exhibits a high peak at zero and a heavy tail, consistent with the notion that complex cell responses to natural images are sparse (Hyvärinen and Hoyer, 2000).

2.2 Sparse coding of complex cell responses

Having sampled a large number of activation patterns (such as those shown in Figure 5b), we then trained a linear sparse coding network on this data. In other words, we estimated the parameters of the model represented by equation (2) and depicted as a network in Figure 3. (The details of the learning procedure are given in the Appendix.) Each complex cell activation pattern gives one data vector $\mathbf{x}$, with each element x_j representing the firing rate of one neuron. Each s_i represents the response of one higher-order neuron, whose 'receptive field' is closely related to the corresponding $\mathbf{a}_i$. Again, the goal is to find basis patterns $\mathbf{a}_i$ such that typical input patterns $\mathbf{x}$ can be described accurately using only a few significantly active higher-order neurons. This kind of *sparse coding of complex cell responses* is illustrated in Figure 6.

Because our input data (complex cell responses) cannot go negative it is natural to require our generative

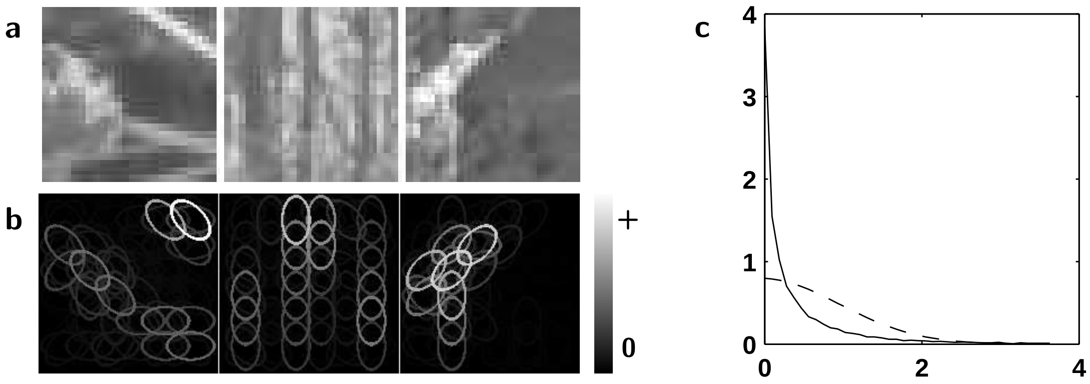

Figure 5: Model complex cell responses to natural image patches. **(a)** Three patches from the set of natural images. **(b)** Responses of the model complex cells to the patches. The ellipses show the orientation and approximate extent of the individual complex cells. The brightness of the different ellipses indicate the response strengths. **(c)** Response distribution of a single complex cell. The solid line shows the normalized histogram of the response of a single complex cell, measured over all image patches. For comparison, the dotted line is the density of the absolute value of a Gaussian random variable. The distributions have been normalized to the same scale (as measured by the expected squared value of the random variable).

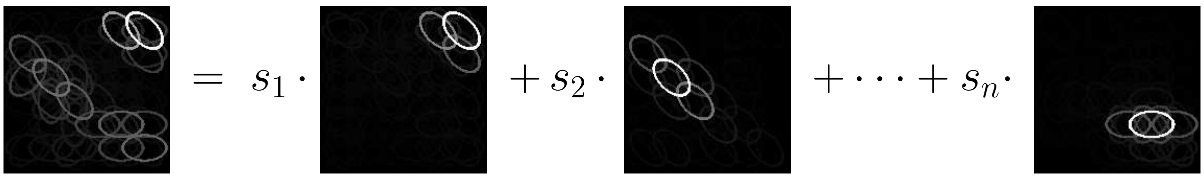

Figure 6: Sparse coding of complex cell responses. Each complex cell activity pattern is represented as a linear combination of basis patterns $\mathbf{a}_i$. The goal is to find basis patterns such that the coefficients s_i are as 'sparse' as possible, meaning that for most input patterns only a few of them are needed to represent the pattern accurately. Cf. Equation 2.

representation to be non-negative. Thus both the $\mathbf{a}_i$ and the s_i were restricted to non-negative values.[d] Arguments for non-negative representations (Paatero and Tapper, 1994) have previously been presented by Lee and Seung (1999). However, in contrast to their approach, we emphasize the importance of sparseness in addition to good reconstruction. Such emphasis on sparseness has previously been forcefully argued for by Barlow (1972) and Field (1994). Thus, we combine sparse coding and the constraint of non-negativity into a single model. Note that neither assumption is arbitary; both follow from the properties of complex cell responses, which are sparse and non-negative.

3 Results

3.1 Properties of the learned representation

Using simulated complex cell responses to natural images as input data (see Figure 5), we thus estimated the non-negative sparse coding model, obtaining 288 basis (activity) patterns. A representative subset of the estimated basis patterns $\mathbf{a}_i$ is shown in Figure 7. Note that most basis patterns consist of a variable number of active complex cells arranged collinearly. This makes intuitive sense, as collinearity is a strong feature of the visual world (Krüger, 1998; Sigman et al., 2001; Geisler et al., 2001). In addition, analyzing images in terms of smooth contours is supported by evidence from both psychophysics (Field et al., 1993; Polat and Sagi, 1993) and physiology (Kapadia et al., 1995; Polat et al., 1998; Kapadia et al., 2000), and is incorporated in many models of contour integration, see e.g. (Grossberg and Mingolla, 1985; Li, 1999; Neumann and Sepp, 1999). To our knowledge, ours is the first model to learn this type of a representation from the statistics of natural images.

It is easy to understand why basis patterns consist of collinear complex cell activity patterns: Such patterns are typical in the data set, and can be sparsely coded if a long contour can be represented by only a few higher-level units. The necessity for *different* length basis patterns comes from the fact that long basis patterns simply cannot code short (or curved) contours, and short basis patterns are inefficient at representing long, straight contours. This kind of contour coding is illustrated in Figure 8.

It is, however, not obvious that the higher-order representation should necessarily code for contours. Multi-layer mechanisms similar to the one used here have been proposed in the context of texture segregation as well (Sperling, 1989; Malik and Perona, 1990). A priori, one could have expected such texture boundary detectors to emerge from our model. Our results seem to indicate that contour coding is, at least in this sparse coding sense, more basic than texture segregation. One must note, however, that we used only a single spatial scale whereas texture segregation could be much more efficient when several spatial frequencies are used.

To characterize the whole population of basis patterns, we described each pattern in terms of 5 parameters: location (x and y), orientation, length, and width (see Appendix for details). We then investigated how the basis vectors are distributed in this parameter space. The main results are shown in Figure 9. First, note that the positions of the basis pattern are distributed relatively evenly inside the window (Fig. 9a). Fig. 9b shows that cardinal orientations are represented slightly better than oblique ones. This could be due partly to a similar bias in natural images (Coppola et al., 1998). Another possibility is that the rectangular complex cell sampling array biases these results, as distances between collinear complex cells are longer for oblique than cardinal orientations.

The most interesting result, however, is that of the distribution of pattern lengths (Fig. 9c).[e] The histogram makes clear that short basis patterns are most abundant, with longer ones progressively more scarce. This is reminiscent of the spatial frequency distribution of a wavelet filter set, where the number of filters needs to increase quadratically with the spatial frequency in order to yield a scale-invariant

[d] Allowing either $\mathbf{a}_i$ or s_i to take negative values would imply that our model would assign non-zero probability density to negative $\mathbf{x}$. As our data is non-negative it therefore makes sense to require the same of both the basis and the coefficients. In our experiments, however, we noted that the non-negativity constraint on the $\mathbf{a}_i$ was not strictly necessary, as the $\mathbf{a}_i$ tended to be positive even without the constraint. Rather, it is the constraint on the s_i that is crucial.

[e] The width parameter did not vary much, since practically all basis patterns consisted of only one row of collinearly active complex cells. Thus, the length parameter is for all practical purposes equal to pattern elongation (length/width).

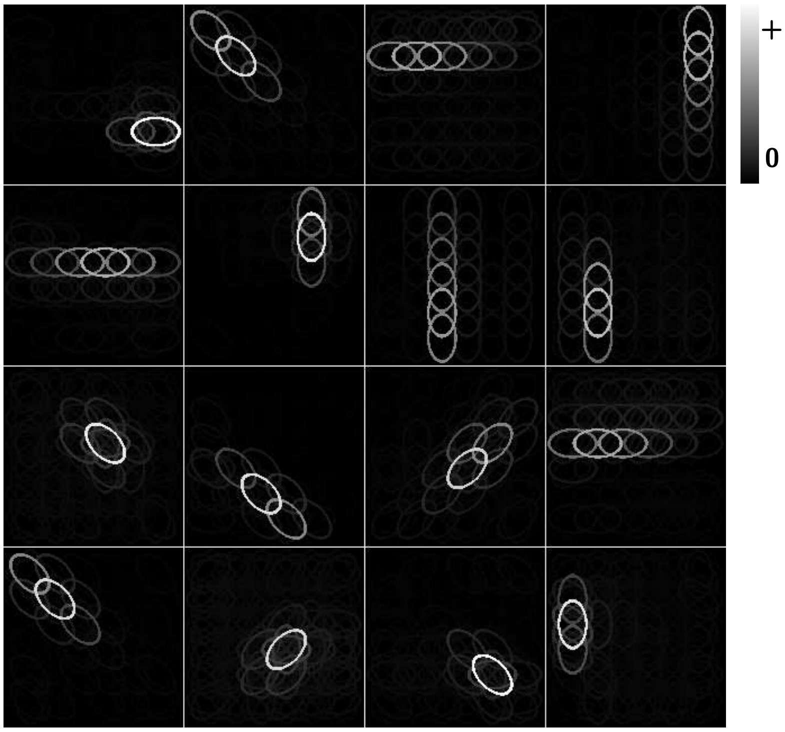

Figure 7: A representative set of basis functions from the learned basis. The majority of units code the simultaneous activation of collinear complex cells, indicating a smooth contour in the image.

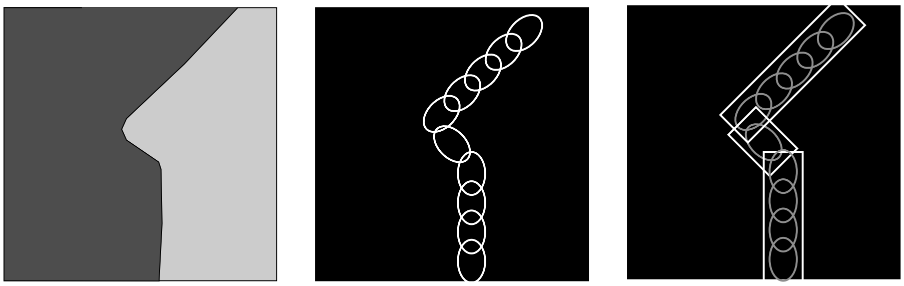

Figure 8: Contour coding in the model. A hypothetical contour in an image (left) is transformed into complex cell responses (middle). These responses can be sparsely represented using only three higher-order units (right) of the types shown in Figure 7.

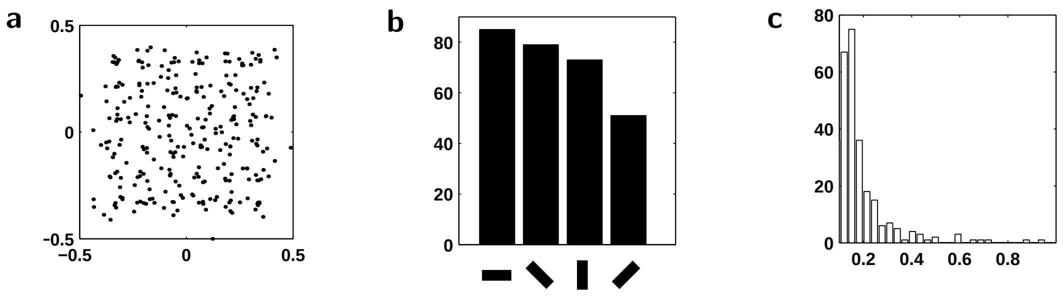

Figure 9: Characterizing the whole population of basis patterns. **(a)** Locations of basis patterns. Each dot indicates the (central) position of one basis pattern in the sampling window. **(b)** Histogram of basis pattern orientation. **(c)** Distribution of pattern lengths, in units relative to the width of the sampling window.

representation (Olshausen and Field, 1997).

An important question is how well these results can be explained simply as resulting from the linear correlations between complex cell outputs (as opposed to higher-order statistics). In Figure 10a we show the covariance structure of the data. The top pattern shows the covariance of one unit (the brightest ellipse) of a cardinal orientation with all other units. The bottom patterns gives the corresponding pattern for a complex cell of an oblique orientation. The strongest correlations are between collinear units, but parallel units also show relatively strong correlations. These observations are compatible with results from previous studies (Krüger, 1998; Sigman et al., 2001; Geisler et al., 2001). Note, however, that we found no clear co-circularity effect (Sigman et al., 2001; Geisler et al., 2001) in our data.

The fact that the learned basis patterns (Figure 7) show much stronger collinearity (as compared with parallel structure) than present in the data covariance is one indication that something beyond covariance structure has been found. However, much stronger evidence comes from the fact that learned patterns have highly varying lengths. This is a non-trivial finding that cannot be explained simply from the linear correlations in the data. The fundamental principle behind this feature of the learned decomposition is sparseness. In fact, if we omit the sparseness objective and simply optimize reconstruction under the non-negativity constraint (i.e. we perform non-negative matrix factorization with the squared error objective, using the algorithm in (Lee and Seung, 2001)) we do get collinearity but not any significant variation in basis pattern lengths. (For illustration, a few such basis patterns are shown in Figure 10b.) We will return to the length distribution in Section 3.2.

A second question is whether the dependencies observed between our model complex cells could be to a significant degree due to the particular choice of forward transform (the complex cell model and the chosen sampling grid) and not natural image statistics. To investigate this we fed our network with image patches consisting of Gaussian white noise, and examined the learned basis patterns; a small subset is shown in Figure 10c. These basis patterns exhibited only a very low degree of collinearity and were not localized inside the patch. This shows clearly that our results are to the most part a consequence of the input statistics and not simply due to the particular complex cell transform chosen.

3.2 Nonlinear responses of higher-order neurons

Although the network is linear from the latent variables s_i to the data $\mathbf{x}$, the inferred (most likely) s_i are a nonlinear function of the data $\mathbf{x}$, due to the noise and the *overcompleteness* of the basis (Olshausen and Field, 1997). In other words, the contour-coding neurons respond to the complex cell activity patterns in a nonlinear fashion. In particular, there is competition between the neurons (Olshausen and Field, 1997), so that they respond only when they are better than competing units at representing the stimulus. As a prominent feature of the found representation is the existence of different-length patterns, this leads to

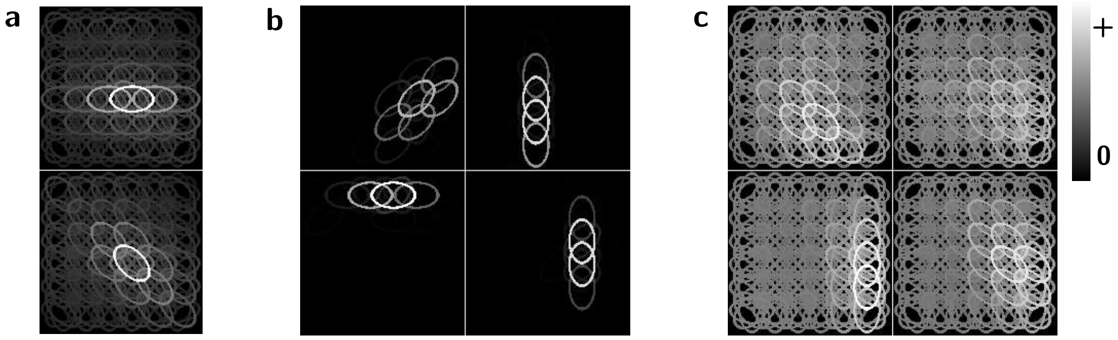

Figure 10: Various additional experiments. **(a)** Covariance structure of complex cell responses. The brightness of each ellipse shows the covariance of that complex cell with the one represented by the brightest ellipse. Top: Covariance of all units with a complex cell of a cardinal orientation. Bottom: Covariance of all units with a cell exhibiting an oblique orientation. **(b)** A few representative basis patterns resulting from applying non-negative matrix factorization to the data. This is equivalent to our model but without the sparseness objective. See main text for discussion. **(c)** Typical basis patterns when our model is learned from image patches consisting of white noise.

units being selective for contour length, in addition to being tuned to position and orientation. In other words, units representing long contours do not respond to short ones, whereas units coding short contours exhibit *end-stopping* (Hubel and Wiesel, 1965; Hubel and Wiesel, 1968).

To illustrate the nonlinear transform from complex cell activities $\mathbf{x}$ to higher-order activities s_i we can make a linear approximation (for details, see the Appendix). Optimal approximating linear filters are shown in Figure 11b, for the units whose basis patterns are depicted in Figure 11a. Note that units representing short contour segments tend to have inhibitory regions at one (or both) of the ends of their 'receptive fields', illustrating the end-stopping effect. On the other hand, units which code longer contours have inhibitory weights from complex cells which are positioned on the contour but are of the wrong orientation.[f] This enhances the selectivity of these units so that they don't respond to contours that only partly overlap the receptive field.

The nonlinear effects can also be seen by directly showing length-tuning curves (Figure 11c). Each plot shows how the response of the corresponding higher-order neuron varies with the length of the stimulus, when all other stimulus parameters are held at their optimal values. The length of the stimulus (relative to the length of the sampling window) is given on the horizontal axis (note the logarithmic scale) and the corresponding response is plotted on the vertical axis. Notice how the response of the end-stopped units starts to decrease when the stimulus length increases past its optimal value, eventually falling to zero. On the other hand, the response of the unit coding long contours does not decline by any significant degree. These results thus show that our model higher-level units have extra-classical properties that make them clearly distinct from standard complex cells.

It should be noted that those higher-order neurons which represent long contours bear many similarities to 'collator' (or 'collector') units, proposed in the psychophysical litterature (Mussap and Levi, 1996; Moulden, 1994). Such units are thought to integrate the responses of smaller, collinear filters, to give a more robust estimate of global orientation than could be achieved with elongated linear mechanisms.[g]

[f]This is a bit difficult to see from the figure, as the negative weights (dark ellipses) are partly masked by the strong positive weights (bright ellipses).

[g]In principle, long contours could be represented by long basis vectors on the level of simple cells as well. However, the representation by these higher-order contour coding cells has the advantage of being less sensitive to small curvature and other departures from strict collinearity. Even very small curvature can completely change the response of an elongated linear filter (simple cell), but it does not change the representation on this higher level, assuming that the curvature is so small that the line stays inside the receptive fields of the same complex cells. Thus, higher-order contour cells give a more robust representation of the contours.

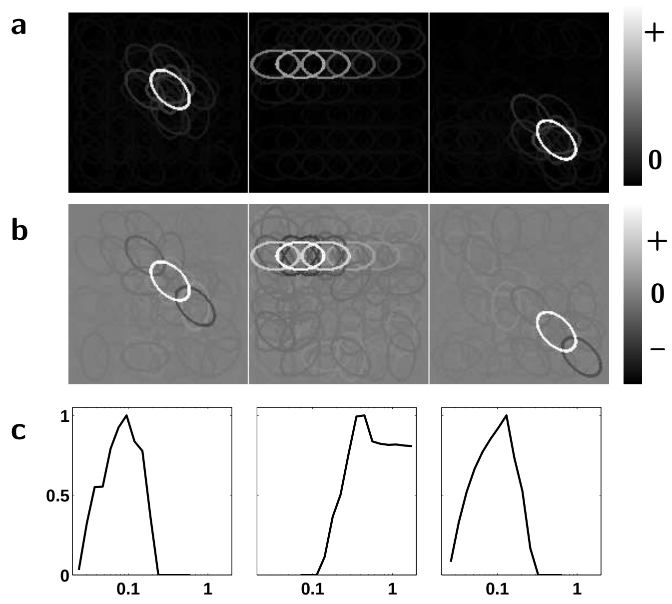

Figure 11: **(a)** Three basis patterns from the estimated basis. **(b)** Optimal approximating linear filters for the units in (a). These are the filters that minimize the mean squared error between the linear response (followed by half-rectification) and the optimal activations. **(c)** Length-tuning curves for the units in (a). The horizontal axis gives the length of the stimulus (logarithmic scale, relative to the size of the sampling window) and the vertical axis denotes response strength (normalized to a maximum of one).

3.3 Contour integration viewed as top-down inference in the model

A central question in visual neuroscience concerns the computational role of feedback connections. It has been suggested that the purpose of feedback is that of using information from higher-order units to modulate lower-level outputs, so as to selectively enhance responses which are consistent with the broader visual context (Lamme, 1995). In hierarchical generative models, this is naturally understood as part of the inference process: Finding the most likely configuration of the network requires integrating bottom-up evidence with top-down priors at each layer of the network (Mumford, 1994; Hinton and Ghahramani, 1997).

Why would this kind of feedback inference be useful? In many cases, there can be multiple conflicting interpretations of the stimulus even on the lowest level, and top-down feedback is needed to resolve such conflicts. In essence, feedback inference computes the most likely interpretation of the scene (Knill and Richards, 1996), combining bottom-up sensory information with top-down priors.

In particular, we suggest that contour integration could be viewed as a natural consequence of such inference in a full hierarchical model. Basically, the argument goes as follows: If enough collinear complex cells are active, they will activate a higher-order contour-coding unit. The activation of such a unit is then evidence for a contour at that location, and this evidence will strengthen responses of all complex cells lying on the contour, especially those whose bottom-up input is relatively weak.

In our simplified network model (Figure 3), the responses of the complex cells are a straightforward function of the image input (through the energy model). How can we then simulate the full network inference process described above? In the unrestricted network, the top-down connections from the contour-coding units to the complex cells seek to adjust the complex cell responses towards that predicted by the contour units:

$$\tilde{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{a}_i s_i, \tag{3}$$

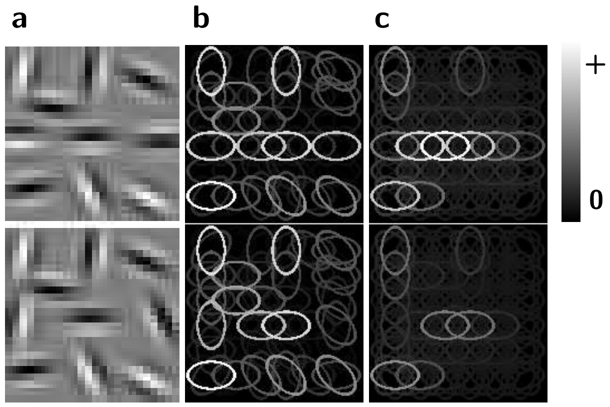

Figure 12: Noise reduction and contour integration. **(a)** Two image patches containing Gabors at random locations and orientations. In the top patch there is a collinear set of three Gabors, whereas in the bottom patch these same Gabors had random orientations. **(b)** The response of the model complex cells to the images in (a). **(c)** The response of the complex cells after feedback noise reduction using the learned network model. Note that the reduction of noise has left the activations of the collinear stimuli but suppressed activity that did not well fit the learned sparse coding model.

Note that this is essentially Equation 2 with the noise term dropped. In our simplified model, this can be seen as a *reduction of noise* in the bottom-up signals (Hupé et al., 1998). (Note that 'noise' in this context refers to any activity that is not consistent with the learned statistical model and is thus not only neural or dark noise.) Such noise-reduction (Hyvärinen, 1999b; Lewicki and Olshausen, 1999; Simoncelli and Adelson, 1996) essentially suppresses responses which are not typical of the training data, while retaining responses that do fit the learned statistical model.

In Figure 12, we show a very basic example of how feedback noise reduction in our model results in the emphasis of smooth contours. We generated image patches by placing Gabor filters at random locations and orientations. In one case, there was a collinear alignment of three consecutive Gabors; in the other these same Gabors had random orientations. These image patches are shown in Figure 12a. Next, we processed these by our model complex cells, as we had processed the natural image patches in our main experiments. The result is shown in Figure 12b. Finally, we calculated the contour-coding unit activities s_i, and plotted the noise-reduced complex cell activity $\tilde{\mathbf{x}} = \sum_{i=1}^{n} \mathbf{a}_i s_i$ in Figure 12c.

Note how the noise reduction step supresses responses to 'spurious' edges, while emphasizing the responses that are part of the collinear arrangement. Such response enhancement to contours is the defining characteristic of many proposed computational models of contour integration, see for example (Grossberg and Mingolla, 1985; Li, 1999; Neumann and Sepp, 1999). Comparing the denoised responses (Figure 12c) with each other one can also observe collinear contextual interactions in the model. The response to the central gabor is stronger when it is flanked by collinear gabors (upper row) than when the flankers have random orientations (bottom row), even though the flankers fall well outside the receptive field of the central neuron. This type of contextual interactions have been the subject of much study recently (see e.g. (Kapadia et al., 1995; Polat et al., 1998; Kapadia et al., 2000)), and are hypothesized to be related to contour integration, although such a relation is not certain.

Although this experiment is exceedingly simple, it nevertheless suggests that something like contour integration would be a natural result of inference or noise reduction in a network like ours. Further experiments will be needed to examine how this noise reduction mechanism would perform on contours in natural images. These are, however, left to future work.

4 Discussion

4.1 Related work

Several authors (Krüger, 1998; Sigman et al., 2001; Geisler et al., 2001) have studied the mutual dependencies of 'complex cell' responses from natural images. The main result from these studies is that complex cell responses are not mutually independent, and in particular they are correlated when the complex cells are arranged collinearly, in parallel, or on a common circle (Krüger, 1998; Sigman et al., 2001; Geisler et al., 2001). However, these investigations only considered pairwise statistics of responses, not higher-order statistics. From previous work in the area of natural image statistics, we know that higher-order statistics are of fundamental importance, for example in explaining oriented receptive fields (Field, 1994; Olshausen and Field, 1996; Zetzsche and Krieger, 1999). Thus it is probably important to consider more than linear correlations also when analyzing dependencies between complex cell outputs. As explained in Section 3.1, the finding that a sparse code of complex cell responses employs basis patterns of varying lengths is a clear indication that something above covariance structure has been learnt.

A further difference to previous work is that in our generative model collinearity is the dominant property and parallel or co-circular basis patterns were not found. The lack of co-circular basis vectors might of course be caused by the limited sampling in space and orientation by our model complex cells, which thus cannot represent all curves equally well. However, preliminary larger-scale experiments (with tighter complex-cell sampling in space and in orientation) seemed to indicate that this was not the case. Rather, it is possible that collinearity is special in the sense that only straight contours are coded by 'collator' units while curved contours are best represented by standard complex cells (possibly bound by horizontal connections). Or, perhaps curved contours should really be represented at the next hierarchical layer: In such a representation, curved contours would be described as the activation of a suitable combination of end-stopped units such as those found in this study.

An approach related to ours is given by the predictive coding model of Rao and Ballard (1999). They proposed a multi-layer network where the higher-level units provide predictive feedback to the lower units. The lower-level neurons then send forward residuals between the observed data (the stimulus) and the prediction. There are no fixed weights, thus their network resembles the full network model (discussed in Section 1.3) although the generative model is very different (due to a difference in the form of the nonlinearities). The main advantage of our approach is that the results can be interpreted intuitively in terms of contour coding (collator) units. Another major difference in their work lies in how they interpreted extra-classical effects such as end-stopping. They proposed that end-stopping was due to cells sending residuals between top-down predictions and bottom-up input: an end-stopping cell stopped responding when the stimulus length was increased because then it could be predicted and there was no residual. However, we interpret end-stopping as a property arising from a competition to represent the stimulus between neurons representing different-length stimuli.

Another recent study related to ours is that given by Park et al. (2000). They show how receptive fields sensitive to translation, contraction, expansion, and rotation (such as those observed in area MSTd) emerge when learning a sparse code for optical flow fields. Although the inputs to their model are vector fields, not simulated neuron outputs, their work nevertheless shows that even quite complex neuronal response properties could be learned from the visual input statistics in a completely unsupervised manner.

4.2 Possible neural implementation

So far, we have not said much about how we hypothesize this kind of a representation to be implemented in the biological visual system. There are several issues that need to be considered.

First, note that although we use a nonlinear optimization technique to calculate optimal higher-level activities s_i from the input data $\mathbf{x}$, we have not said anything about how a real neural network might perform this operation. On the basis of arguments relating to processing speed (Thorpe et al., 1996), it would be expected that feedforward processing would be tuned to already give a fairly good estimate of the optimal responses. Such processing might then be complemented by influences through recurrent connections. For a concrete suggestion of how this inference could be performed by a network, see (Olshausen and Field, 1997).

Also note that although contour integration in our model is explicitly handled by feedback connections from higher-order units, this does not rule out the role of horizontal connections between complex cells in V1. On the contrary, in a neural implementation, it would probably be quite useful to implement at least part of the inference/noise reduction by the use of horizontal connections. The contour integration performed by these connections might be viewed as inference or noise reduction as well, thus implicating a possible general purpose of horizontal connections in the visual cortex.

One issue we have also left untouched is the question of where our contour-coding neurons would reside in the cortex. A naïve guess would be area V2. Indeed, V2 is strongly reciprocally connected with V1, and end-stopping (one important property of many of our contour units) seems to be more prevalent there, particularly in the pale cytochrome oxidase stripes (Hubel and Livingstone, 1987). However, end-stopping neurons are also found in V1 (Hubel and Wiesel, 1968), so it is quite possible that part (or all) of our hypothesized contour-coding neurons might be situated there. On the other hand, it is also possible that such units do not exist at all and the horizontal connections between V1 complex cells are utilized to perform the relevant computations. If that turned out to be the case, the contribution of this paper has mainly been to illustrate the existing statistical dependencies and to hint at how those dependencies might be related to contour integration.

4.3 Conclusions

We have extended the sparse coding framework to work on the outputs of model complex cells with natural image input. This provides a way towards predicting and understanding neuronal response properties higher in the visual processing hierarchy. The framework also suggests to view feedback pathways as an essential part of the inference process, performing something like noise reduction on the lower level activities. Specifically, we have shown how contour coding and end-stopped receptive fields can be understood in this sparse coding framework. Furthermore, contour integration might in this framework be interpreted as a natural result of inference in a hierarchical network.

Appendix

To facilitate the reproduction of the results we provide the complete MATLAB code for all experiments at `http://www.cis.hut.fi/phoyer/contournet.tar.gz`.

A sample of 100,000 image patches of 24×24 pixels were sampled from natural images available on the WWW (`http://www.cis.hut.fi/projects/ica/data/images`). The output of a model complex cell was given by

$$C_{\{x_c,y_c,\theta\}} = \left(\sum_{x,y} G_{\{e,x_c,y_c,\theta\}}(x,y) I(x,y) \right)^2 + \left(\sum_{x,y} G_{\{o,x_c,y_c,\theta\}}(x,y) I(x,y) \right)^2 \tag{4}$$

when fed with the image patch $I(x,y)$. Here $G_{\{e,x_c,y_c,\theta\}}$ and $G_{\{o,x_c,y_c,\theta\}}$ are even- and odd-symmetric (respectively) Gabor filters, centered on (x_c, y_c) and of orientation θ. The Gaussian envelope had an aspect ratio of 1.5 and the spatial frequency of the sinusoid was chosen to give $2-3$ main excitatory/inhibitory regions, in accordance with measurements of receptive fields in V1 (DeAngelis et al., 1993). As there were a total of $6 \times 6 \times 4 = 144$ complex cells, this was the dimensionality of our data $\mathbf{x}$.

The model (Eq. 2) was estimated by maximizing the joint log-posterior of the latent variables and model parameters, given the data. This is equivalent to minimizing

$$C(\mathbf{A}, \mathbf{s}^{(n)}, n = 1 \ldots N) = \sum_n \left[\|\mathbf{x}^{(n)} - \mathbf{A}\mathbf{s}^{(n)}\|^2 + \lambda \sum_i s_i^{(n)} \right] \tag{5}$$

where $\mathbf{A}$ denotes the matrix containing the $\mathbf{a}_i$ as its columns, $\mathbf{x}^{(n)}$ represents the n:th data vector, and $\mathbf{s}^{(n)}$ is the vector containing the latent variables $s_i^{(n)}$ corresponding to the n:th data vector. The linear activation penalty term comes from assuming the latent variables to be exponentially distributed, i.e. $p(s_i) = \exp(-s_i)$, while the quadratic term derives from the Gaussian prior on the noise. The constant λ defines the tradeoff between representation error and sparseness, and is obtained directly from the noise level (selected manually) assumed in the model. The objective was minimized by standard gradient descent (restricted to non-negative $\mathbf{A}$ and $\mathbf{s}^{(n)}$) with respect to the $\mathbf{s}^{(n)}$ in the short run and with respect to $\mathbf{A}$ under a longer timescale, as in (Olshausen and Field, 1996).[h] An additional constraint was that each latent variable was assumed to contribute equally to the data, i.e. $\|\mathbf{a}_i\| = \|\mathbf{a}_j\|$ for all i, j.

Basis patterns were parametrized by five parameters, location (x and y), orientation, length, and width. The patterns were approximated by a Gaussian cluster of activity of complex cells of the given orientation, at the specified position, and having standard deviations *length*/2 and *width*/2. The parameters were chosen to minimize the summed squared difference between basis patterns and parametrized patterns. This yielded an excellent fit in almost all cases.

The weights shown in Figure 11b were obtained by first finding the optimal s_i for all data points by minimizing the objective (Eq. 5) and then iteratively seeking weights $\mathbf{w}_i$ that minimize $\sum_n \{(s_i^{(n)} - \mathrm{rect}(\mathbf{w}_i^T \mathbf{x}^{(n)}))^2\}$, where rect() denotes half-rectification.

The noise reduction was done by, for a given data vector $\mathbf{x}$, finding the most likely latent variable configuration by minimizing $\|\mathbf{x} - \mathbf{A}\mathbf{s}\|^2 + \lambda \sum_i s_i$ with respect to $\mathbf{s}$. The denoised data vector is then given by $\mathbf{A}\mathbf{s}$.

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.

Atick, J. J. and Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, 4:196–210.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61:183–193.

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In Rosenblith, W. A., editor, *Sensory Communication*, pages 217–234. MIT Press.

Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1:371–394.

Barlow, H. B. (2001a). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences*, 24.

Barlow, H. B. (2001b). Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–253.

Bell, A. J. and Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research*, 37:3327–3338.

Coppola, D. M., Purves, H. R., McCoy, A. N., and Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Science, USA*, 95:4002–4006.

Dan, Y., Atick, J. J., and Reid, R. C. (1996). Efficient coding of natural scences in the lateral geniculate nucleus: experimental test of a computational theory. *Journal of Neuroscience*, 16:3351–3362.

Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, 7:889–904.

[h]From the viewpoint of model estimation, the non-negativity constraint on s_i is important because it allows us to use a more accurate model of the prior densities of the s_i. The main feature of the density is a sharp peak at zero. Most ICA estimation methods assume that this sharp peaks coincides with the mean of the distribution, which is not at all the case here due to the strong asymmetry of the nonnegative distribution. This is why conventional ICA algorithms (Hyvärinen, 1999a) are not able to successfully estimate the components for real-life noisy data characterized by nonnegativity. Note that Parra et al. (2000) also considered the constraint of non-negativity, but their constraint was only on the basis vectors $\mathbf{a}_i$, not on the latent variables s_i.

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69:1091–1117.

Dong, D. W. and Atick, J. J. (1995). Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, 6:159–178.

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

Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local 'association field'. *Vision Research*, 33(2):173–193.

Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge co-occurence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724.

Grossberg, S. and Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries and emergent segmentations. *Perception and psychophysics*, 38(2):141–171.

Hinton, G. E., Dayan, P., Frey, B. J., and Neal, R. M. (1995). The wake-sleep algorithm for unsupervised neural networks. *Science*, 268:1158–1161.

Hinton, G. E. and Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations. *Phil. Trans. R. Soc. Lond. B*, 352:1177–1190.

Hoyer, P. O. and Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems*, 11(3):191–210.

Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154.

Hubel, D. and Wiesel, T. (1965). Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28:229–289.

Hubel, D. and Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195:215–243.

Hubel, D. H. and Livingstone, M. S. (1987). Segregation of form, color, and stereopsis in primate area 18. *Journal of Neuroscience*, 7:3378–3415.

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394:784–787.

Hyvärinen, A. (1999a). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10(3):626–634.

Hyvärinen, A. (1999b). Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11(7):1739–1768.

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(7):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(18):2413–2423.

Hyvärinen, A., Hoyer, P. O., and Inki, M. (2001). Topographic independent component analysis. *Neural Computation*, 13(7):1527–1558.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley Interscience.

Kapadia, M. K., Ito, M., Gilbert, C. D., and Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, 15(4):843–856.

Kapadia, M. K., Westheimer, G., and Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, 84:2048–2062.

Knill, D. C. and Richards, W., editors (1996). *Perception as Bayesian Inference*. Cambridge University Press.

Krüger, N. (1998). Collinearity and parallelism are statistically significant second order relations of complex cell responses. *Neural Processing Letters*, 8:117–129.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, 15(2):1605–1615.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing 13 (Proc. NIPS*2000)*. MIT Press.

Lewicki, M. and Olshausen, B. (1999). Probabilistic framework for the adaption and comparison of image codes. *J. Opt. Soc. Am. A: Optics, Image Science, and Vision*, 16(7):1587–1601.

Li, Z. (1999). Pre-attentive segmentation in the primary visual cortex. *Spatial Vision*, 13(1):25–50.

Malik, J. and Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7(5):923–932.

Morrone, M. C. and Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proc. Royal Soc. London Ser. B*, 235(1280):221–245.

Moulden, B. (1994). Collator units: Second-stage orientational filters. In *Higher-order processing in the visual system. Ciba Foundation Symposium 184*, pages 170–192. John Wiley, Chichester, U.K.

Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In Koch, C. and Davis, J., editors, *Large-scale neuronal theories of the brain*. MIT Press, Cambridge, MA.

Mussap, A. J. and Levi, D. M. (1996). Spatial properties of filters underlying vernier acuity revealed by masking: Evidence for collator mechanisms. *Vision Research*, 36(16):2459–2473.

Neumann, H. and Sepp, W. (1999). Recurrent V1-V2 interaction in early visual boundary processing. *Biological Cybernetics*, 81:425–444.

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

Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325.

Paatero, P. and Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126.

Park, K.-Y., Jabri, M., Lee, S.-Y., and Sejnowski, T. (2000). Independent components of optical flows have MSTd-like receptive fields. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pages 597–602.

Parra, L., Spence, C., Sajda, P., Ziehe, A., and Müller, K.-R. (2000). Unmixing hyperspectral data. In *Advances in Neural Information Processing 12 (Proc. NIPS*99)*, pages 942–948. MIT Press.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., and Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, 391:580–584.

Polat, U. and Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33:993–999.

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

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience*, 2(1):79–87.

Sigman, M., Cecchi, G. A., Gilbert, C. D., and Magnasco, M. O. (2001). On a common circle: Natural scenes and gestalt rules. *Proceedings of the National Academy of Science, USA*, 98:1935–1940.

Simoncelli, E. P. and Adelson, E. H. (1996). Noise removal via bayesian wavelet coring. In *Proc. Third IEEE Int. Conf. on Image Processing*, pages 379–382, Lausanne, Switzerland.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1215.

Simoncelli, E. P. and Schwartz, O. (1999). Modeling surround suppression in V1 neurons with a statistically-derived normalization model. In *Advances in Neural Information Processing Systems 11*, pages 153–159. MIT Press.

Sperling, G. (1989). Three stages and two systems of visual processing. *Spatial Vision*, 4(2/3):183–207.

Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proc. Royal Society ser. B*, 216:427–459.

Tailor, D. R., Finkel, L. H., and Buchsbaum, G. (2000). Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research*, 40:2671–2676.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381:520–522.

van Hateren, J. H. and Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. *Proc. Royal Society ser. B*, 265:2315–2320.

van Hateren, J. H. and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal Society ser. B*, 265:359–366.

Wachtler, T., Lee, T.-W., and Sejnowski, T. J. (2001). Chromatic structure of natural scenes. *Journal of the Optical Society of America A*, 18(1):65–77.

Zetzsche, C. and Krieger, G. (1999). Nonlinear neurons and higher-order statistics: new approaches to human vision and electronic image processing. *Proc. SPIE*, 3644:2–33.

Acknowledgements

The authors wish to thank Bruno Olshausen for helpful discussions and for providing the sparsenet source code (on which our code is based). In addition, thanks are due to Jarmo Hurri, Mika Inki, Pentti Laurinen, Zhaoping Li and Harri Valpola for useful suggestions and valuable comments on the manuscript, Jaakko Väyrynen for performing additional experiments, and Erkki Oja for a stimulating work environment. Finally, we are grateful to the two anonymous referees who pointed out several weaknesses in the first draft of the manuscript and thus markedly helped improve the quality of the paper. P.O.H. was funded by the Helsinki Graduate School in Computer Science and Engineering. A.H. was funded by the Academy of Finland.