

# Development of Feature Detectors by Self-Organization

## A Network Model

J. Rubner<sup>1</sup> and K. Schulten<sup>2</sup>

<sup>1</sup> Physik-Department, Technische Universität München, D-8046 Garching, Federal Republic of Germany

<sup>2</sup> Department of Physics, University of Illinois, Urbana, IL 61801, USA

**Abstract.** We present a two-layered network of linear neurons that organizes itself as to extract the complete information contained in a set of presented patterns. The weights between layers obey a Hebbian rule. We propose a local anti-Hebbian rule for lateral, hierarchically organized weights within the output layer. This rule forces the activities of the output units to become uncorrelated and the lateral weights to vanish. The weights between layers converge to the eigenvectors of the covariance matrix of input patterns, i.e., the network performs a principal component analysis, yielding *all* principal components. As a consequence of the proposed learning scheme, the output units become detectors of orthogonal features, similar to ones found in the brain of mammals.

## 1 Introduction

One of the remarkable aspects of our visual system is the fact that we can rapidly discern and identify the important features of our visual environment. Visual information is processed on successive levels of the cortex. Simple features, edges or bars for instance, are detected on the lowest levels, while the recognition of complicated features takes place in higher areas of the visual cortex. A whole range of specialized cells, so-called feature detectors, are responsible for extracting very specific features – like a moving bar – from a visual scene (see, e.g., Hubel and Wiesel 1979). As those feature detectors only process information from a small part of the visual space, they act as filters on a stimulus.

Although part of the synaptic connections in the brain is genetically specified, postnatal visual input plays an essential role in the organization, birth and death of synapses (see, e.g., Cowan 1979). It is unlikely that this organization follows a global plan. One rather

expects that local rules, like Hebb's rule (Hebb 1949), govern the postnatal organization of the brain and the formation of feature detectors. These expectations raise the general question how a sensory system, in response to input information, can organize itself according to local rules so as to form feature detectors which encode mutually independent aspects of the information contained in patterns presented to it.

In the following section, we present a simple, two-layered neural network as a model for such a system. We then test this model, presenting first colors and then spatially varying intensity patterns to the network.

## 2 The Network Model

The proposed network is two-layered, consisting of an input and an output layer with  $N_i$  and  $N_o$  neurons, respectively. The input and output units exhibit real, continuous-valued activities  $\mathbf{i} = (i_1, \dots, i_{N_i})$  and  $\mathbf{o} = (o_1, \dots, o_{N_o})$ . The two layers are completely interconnected and the strength of the synaptic connection between input unit  $j$  and output unit  $m$  is denoted by  $w_{jm}$ . The set of weights connecting an output unit  $m$  with all input units forms the weight vector  $\mathbf{w}_m$ , the transpose of which is the  $m$ -th row of the weight matrix  $\mathbf{W}$ . The set of  $N_\pi$  presented patterns is denoted by  $\{\mathbf{p}^\pi = (p_1^\pi, \dots, p_{N_i}^\pi), \pi = 1, \dots, N_\pi\}$ . Activities of the input units correspond to the presented patterns, i.e.,  $\mathbf{i} = \mathbf{p}^\pi$ . Activities of the output units, in response to a pattern  $\mathbf{p}^\pi$ , are linear sums of the inputs weighted by the synaptic strengths, i.e.,  $\mathbf{o}^\pi = \mathbf{W}\mathbf{p}^\pi$ .

Weights between layers are adjusted upon presentation of an input pattern  $\mathbf{p}^\pi$  according to a Hebbian rule, leading to an increase in synaptic strength if the corresponding pre- and postsynaptic potentials are of the same sign. If weight changes are small enough, the update can be performed after presentation of all

patterns, i.e.,

$$\Delta w_m = \eta \langle (p^\pi - \langle p^\pi \rangle) (o_m^\pi - \langle o_m^\pi \rangle) \rangle, \quad (1)$$

where  $\eta$  is a positive parameter and where the brackets  $\langle \dots \rangle$  denote the average over the set of patterns. The subtraction of averages in (1) can be interpreted as the existence of thresholds of units. On the other hand, subtracting averages is convenient from a mathematical point of view (Oja 1982) and allows us to assume in the following  $\langle p^\pi \rangle = 0$  and  $\langle o^\pi \rangle = 0$ .

Let us first consider the case of one output unit. Linsker (1988) showed that the weights of that unit, if they obey a Hebbian rule like (1), evolve to maximize the variance of the output for a set of presented patterns. If the weights are normalized after every update such that  $\sum_i w_{i1}^2 = 1$ , the Hebbian rule renders

weights which characterize the direction of maximal variance of the pattern set (Oja 1982). Equivalently, the weight vector  $w_1$  converges to the eigenvector with the largest eigenvalue of the covariance matrix  $C$  of the pattern set, with elements  $C_{jk} = \langle p_j^\pi p_k^\pi \rangle$ . Diagonalizing a covariance matrix corresponds to the statistical technique of *principal component analysis* (see, e.g., Lawley and Maxwell 1963). Thus, a Hebbian learning rule for Euclidian normalized weights yields the first principal component of the input data set. The non-vanishing weights  $w_{i1}$  of the output unit define its receptive field. The output unit then corresponds to a feature detector which analyses the principal feature of a presented pattern. In other words, the net represents a so-called matched linear filter (Oja 1982).

However, a single unit only processes a fraction of the total information contained in a pattern. In order to transmit the complete information between the two layers, as many output cells as the rank of the covariance matrix  $C$  are required. Furthermore, in order to become filters of mutually orthogonal features, the output cells need to become uncorrelated. For that purpose, we assume the existence of lateral, hierarchically organized connections with weights  $u_{lm}$  between output units  $l$  and  $m$ , where  $l < m$ . The activity of the  $m$ -th output cell is then given by  $o_m^\pi = w_m \cdot p^\pi + \sum_{l < m} u_{lm} w_l \cdot p^\pi$ . Then, following (1), changes of synaptic connection strengths between input units and output unit  $m$  are given by

$$\Delta w_m = \eta \left( C w_m + \sum_{k < m} u_{km} C w_k \right). \quad (2)$$

Figure 1 shows a scheme of such network.

We propose that lateral weights adapt themselves according to an *anti-Hebbian* rule: the strength of a lateral synapse is lowered, if the corresponding pre- and postsynaptic activities are of the same sign. Again, we assume that changes are small, yielding a decrease

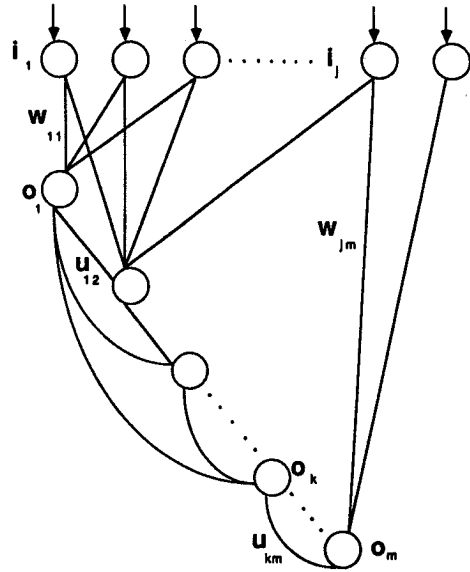


Fig. 1. Scheme of the proposed network. It consists of  $N_i$  input and  $N_o$  output elements. Each input cell  $j$  is connected with each output cell  $m$ , with connection strength  $w_{jm}$ . The weights between output cells are hierarchically organized, with connection strength  $u_{mn}$  between units  $m$  and  $n$ . For the sake of simplicity, only part of the units and connections is shown

in synaptic strength, if the corresponding output units have correlated activities, namely

$$\Delta u_{lm} = -\mu \langle o_l^\pi o_m^\pi \rangle, \quad (3)$$

where  $\mu$  is a positive learning parameter. The anti-Hebbian rule is similar to the learning rule of Kohonen's novelty filter (Kohonen 1982) and to the "unlearning" rule proposed by Hopfield et al. (1983).

As a result of the proposed learning scheme, the weight vector  $w_m$  converges to the eigenvector  $c_m$  of  $C$ . Equivalently, the output units, with initially random receptive fields, become orthogonal feature detectors that analyse mutually independent aspects of the information contained in the set of presented patterns. Because the output cells become uncorrelated, the lateral connections vanish. A detailed mathematical proof of these findings is given elsewhere (Rubner and Tavan 1989).

Several authors have proposed inhibitory connections between output units in order to render their activities uncorrelated (von der Malsburg 1973; Kohonen 1982; Rumelhart and Zipser 1985; Yuille et al. 1989). In our scheme, lateral connections are both excitatory and inhibitory before they vanish, due to the anti-Hebbian rule. This results in a purely feed-forward network, which represents an important computational advantage for a parallel system. Principal component analysis has also been associated with linear feed-forward networks using optimization methods with respect to a quadratic error function, i.e.,

back-propagation (Baldi and Hornik 1989). The advantage of our model consists in optimal feature extraction without supervision and in the existence of biologically plausible, local adaptation rules for the weights.

In the following two sections we illustrate the performance of the proposed network by numerical simulations. One simulation step, or learning cycle, consists in (i) updating weights between layers according to (2), (ii) normalizing these weights, such that  $\sum_i w_{im}^2 = 1$ , and (iii) updating lateral weights according to (3). In applying these rules, it is implied that learning parameters are small enough, so that the weight update can be performed after presentation of all patterns. Initial weights are chosen at random. The asymptotic vanishing of the lateral connection strengths provides a criterion for the convergence of the learning scheme. We choose the number of learning steps large enough that convergence is achieved within numerical precision.

We first present spatially homogeneous color patterns and show that the resulting output cells correspond to color-opponent filters, similar to those found by psychophysical experiments. The last section presents the results of simulations with patterns of spatially varying intensity. The obtained feature detectors are selective to varying orientations and spatial frequencies and their receptive fields are similar to ones of simple cells found in the striate cortex.

### 3 Opponent Color Processing

The first stage of human color processing takes place at the level of the retina, where three different types of color-sensitive photoreceptors are present. They show maximal absorption in the blue, green and red parts of the visible spectrum and, therefore, are called blue, green and red cones. Both psychophysical and neurophysiological experiments have shown that, at deeper lying levels of the retina and subsequent layers of the visual cortex, color processing is done in terms of opponent cone signals. So, psychophysical color matching data have provided evidence for the existence of an achromatic channel and two color-opponent channels, one red-green ( $r - g$ ), the other one blue-yellow ( $b - y$ ) or vice versa (Judd 1949). As far as the neurophysiological aspect is concerned, part of the color sensitive cells in the retina, lateral geniculate nucleus and the visual cortex exhibit spectrally opponent responses to light stimuli (see, e.g., De Valois et al. 1966; Livingstone and Hubel 1984). In order to investigate color processing from an information-theoretical viewpoint, Buchsbaum and Gottschalk (1983) computed the covariance matrix of the cone responses to a set of monochromatic stimuli. A prin-

cipal component analysis of this matrix yields one achromatic and two opponent chromatic channels, similar to the ones recorded by psychophysical experiments. Thus, a transformation from the space of the initial three, partially overlapping cone mechanisms into the space of uncorrelated, opponent channels achieves an efficient information transmission. We therefore expect that the network will organize itself upon presentation of color stimuli so as to form achromatic and color-opponent feature detectors. For this purpose, we have generated a set of 162 different color patterns with varying intensity and chromaticity on an *Apple RGB-monitor*. We have measured the XYZ-values of these color stimuli with a photometer and have computed the responses  $r$ ,  $g$ , and  $b$  of the human cones (Wyszecki and Stiles 1982). Correspondingly, the network consists of three input units, each representing a group of cones and three output units. Initial weights between layers, as well as lateral connection strengths, were chosen at random from the interval  $[-1, +1]$ . The different types of cones are neither equally represented in number nor homogeneously distributed in the retina. Therefore, we take into account effective cone populations, incorporating the effect of spatial inhomogeneities. As red and green cones are morphologically undistinguishable, only neurophysiological data about the distribution of blue cones in the retina are available (Ahnel et al. 1987). The number of blue cones in the central part of the fovea, which is thought to play an essential role in the perception of colors, amounts to about 3% of the total number of cones. The relative number of red and green cones has been estimated by comparing physiologically measured sensitivity curves of red and green cones and the psychophysically determined luminous efficiency function (Schnapf et al. 1987). Following these data, we use for the relative number of red, green and blue cones the ratio of  $N_r : N_g : N_b = 20 : 12 : 1$ . Instead of explicitly taking these ratios into account by simulating 33 input cells, we scale the weights between the red, green and blue input units and the output cells according to the above mentioned ratios.

Figure 2 displays the results of simulations by showing the magnitude of the effective connections leading to the three output units after 5000 learning cycles and for learning parameters  $\eta = \mu = 0.5$ . The strength of scaled synaptic weights from the red, green and blue input cells are drawn on a wavelength axis, at the location of maximal absorption of the red, green and blue cones, respectively. The left part of Fig. 2 shows the weights to the first unit. All three connection strengths are positive, indicating that the first unit computes an intensity measure of the color stimuli. The second unit, with weights displayed in the middle part of Fig. 2 has negative connection strength to the

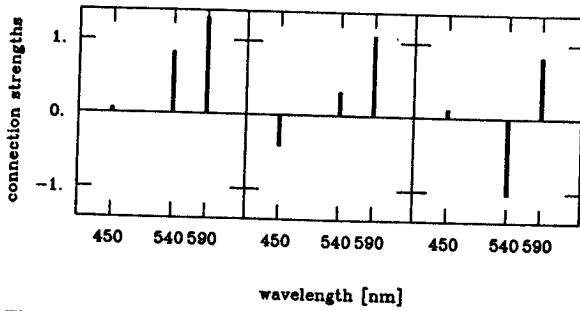


Fig. 2. Scaled synaptic weights between the input units and output units 1-3 (left to right). The connections to an output unit from the red, green and blue input units are drawn on a wavelength scale at the location of maximal absorption of the red, green and blue cones, respectively. The strengths of connections are drawn on an arbitrary scale

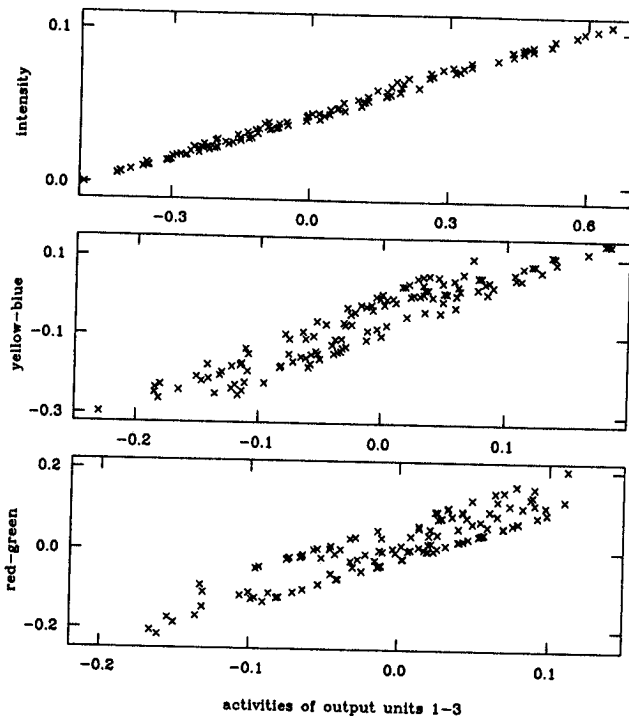


Fig. 3. Psychophysical data of color stimuli (y-axis) versus activities of output units (x-axis). Each cross represents one of the 162 presented color patterns

blue input cell, whereas synaptic weights to the red and green units are positive. Thus, the second output cell resembles a color-opponent feature detector of the  $y-b$  type. The connection from the third output unit (right part of Fig. 2) to the red input unit is excitatory, whereas its connection to the green unit is inhibitory. The third output unit is a  $r-g$  color-opponent cell.

Figure 3 compares these results with psychophysical data, as given by Judd (1949). Each cross represents one color stimulus, where the abscissa is the magnitude of the response of an output cell and the ordinate is the

corresponding psychophysical response. Thus, perfect correlation between the activity of an output cell and a psychophysical channel would cause the crosses to lie on a line. The top part of Fig. 3 correlates the activities of the first output unit in response to a presented pattern with the predicted achromatic channel, while the middle and bottom part correlate the activities of the second and third output units with the predicted  $y-b$  and  $r-g$  channels, respectively. Although the set of 162 presented color stimuli was rather small, all three units are clearly correlated with the psychophysical channels. The fact that the correlation is best for the achromatic unit, is due to the ratio  $N_r:N_g$  of effective red and green cone populations used in the simulation. As mentioned above, this ratio has been obtained by fitting neurophysiological data to the luminous efficiency function, which approximately corresponds to the spectral curve of the achromatic channel. Note, that different ratios  $N_r:N_g:N_b$  yield  $r-g$  and  $b-y$  units with activities that are perfectly correlated with the corresponding psychophysical channels. This indicates that different amounts of red, green and blue cones contribute to the formation of the different color opponent mechanisms.

#### 4 Orientation and Spatial Frequency Selective Cells

The processing of spatial information has, for a long time, been a controversial issue. Simple cells which represent the first stage of spatial information processing in the primary visual cortex were first detected by Hubel and Wiesel, who described them as feature detectors, selective to edges or bars (Hubel and Wiesel 1962). On the other hand, these cells display approximately linear response characteristics and respond to sine-wave gratings. Psychophysical experiments, as well, have revealed the existence of spatial frequency channels (Campbell and Robson 1968; Pollen et al. 1971). There has been a debate over the issue whether the processing of spatial information is based on the extraction of local features or on a Fourier-like decomposition into spatial-frequency channels. However, it appears that both views may be appropriate and even complementary (MacKay 1981). This viewpoint is supported by a successful description of the properties of simple cell receptive fields in a common mathematical framework (Marcelja 1980; Daugman 1985; Jones and Palmer 1987c). In the following, we examine which are the essential features of spatially varying patterns and compare the receptive fields obtained by our learning scheme with the ones of simple cells.

For this purpose, we consider a rectangular lattice of  $N_i \times N_i$  sensory input units representing the receptive field of  $N_o$  output units, with  $N_o \leq N_i$ . We

generate two-dimensional patterns of varying intensity by first selecting random numbers  $s_{ij}^{\pi}$ ,  $\pi = 1, \dots, N_{\pi}$  from the interval  $[-1, +1]$ . Then, in order to introduce information about the topological structure of the receptive field, the random input intensities are correlated, e.g., with their nearest neighbors in both directions. As a result, the component  $p_{ij}^{\pi}$  of a pattern  $\mathbf{p}^{\pi}$  at the coordinate  $(i, j)$  of the receptive field is given by  $p_{ij}^{\pi} = s_{ij}^{\pi} + s_{i-1, j}^{\pi} + s_{i+1, j}^{\pi} + s_{i, j-1}^{\pi} + s_{i, j+1}^{\pi}$ . We assume vanishing boundary conditions, i.e.,  $s_{0j} = s_{i0} = s_{N_i+1, j} = s_{i, N_i+1} = 0$ . Note, that this averaging of neighboring signals corresponds to introducing an additional layer with random activities and with fixed and restricted connections to the input layer.

To outline the emerging filter characteristics of the network, let us first consider the corresponding one-dimensional case. As shown by Rubner and Tavan (1989), the  $j$ -th component of the eigenvector  $\mathbf{w}_k$  of the covariance matrix  $\mathbf{C}$  is proportional to  $\sin[k\pi j/(N_i + 1)]$ . Hence, the eigenvectors are Fourier components of a discrete, finite, one-dimensional chain. Analogously, one can compute the eigenfunctions and eigenvalues of  $\mathbf{C}$  for the two-dimensional case. (Note that, in the case of two-dimensional patterns, the covariance matrix is strictly speaking not a matrix, but a four-dimensional quantity. By combining two indices, we interpret  $\mathbf{C}$  as a  $N_i N_i' \times N_i N_i'$  matrix.) The value of the  $k$ -th eigenfunction at lattice location  $(i, j)$  is proportional to  $\sin[k_1 \pi i/(N_i + 1)] \sin[k_2 \pi j/(N_i' + 1)]$  with integers  $k_1 \in [1, N_i]$  and  $k_2 \in [1, N_i']$ . These eigenfunctions correspond to a two-dimensional sinusoidal plane wave with discretized frequencies, due to the boundary conditions. The eigenvalues are given by

$$t(k_1, k_2) = (1 + 2 \cos[\pi k_1/(N_i + 1)] + 2 \cos[\pi k_2/(N_i' + 1)])^2/3.$$

Figure 4 shows contour plots of the synaptic distribution of the first eight output units after 2000 learning cycles for an input lattice of  $14 \times 10$  units. Solid lines indicate positive, dashed lines negative weights. Learning parameters  $\eta$  and  $\mu$  were set to 0.05 and 0.1, respectively. The displayed receptive fields correspond to the first eight eigenfunctions of the covariance matrix, starting with  $k_1 = k_2 = 1$  (top left part of Fig. 4). The following eigenfunctions appear in order of decreasing eigenvalues. Note that, due to the rectangular lattice, a degeneracy of eigenfunctions, as it would follow from a square input lattice (with  $(k_1, k_2) = (k_2, k_1)$ ), does not occur.

Receptive fields of simple cells in cat striate cortex as recorded by Jones et al. (1987a, b) can be described by Gabor functions (Marcelja 1980; Daugman 1985), which consist of an oscillatory part, namely a sinusoidal plane wave and a Gaussian, exponentially decay-

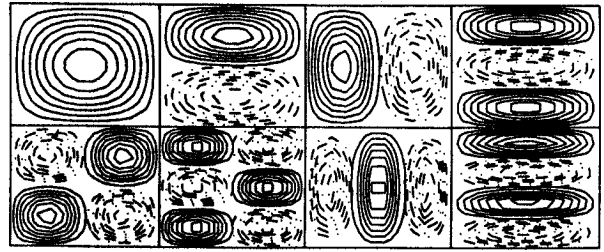
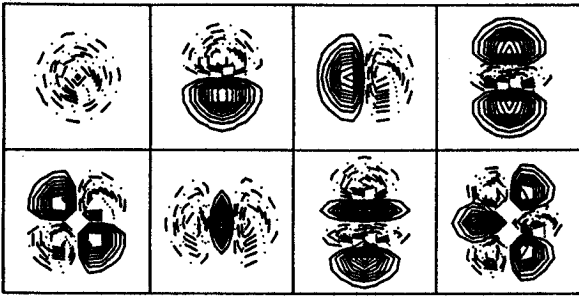


Fig. 4. From left to right and top to bottom: contour plots of receptive fields of output units 1–8 in the case of a rectangular lattice of  $14 \times 10$  input units. The synaptic distribution  $D(i, j)$  between layers was constant. Solid lines indicate positive, dashed lines negative weights. The number of learning cycles was 2000, the learning parameters  $\eta$  and  $\mu$  were equal to 0.05 and 0.1, respectively

ing part. As shown in Fig. 4, our learning scheme yields mutually orthogonal, spatially oscillating receptive fields of various frequencies. However, due to the chosen rectangular geometry of the input layer, an exponential decay is absent. Such a decay can easily be implemented. For that purpose, we scale the weights between layers, i.e., we replace a synaptic weight  $w(ij, m)$  between the input unit at lattice location  $(i, j)$  and the  $m$ -th output unit by  $w'(ij, m) = D(i, j)w(ij, m)$ , where  $D(i, j)$  is a Gaussian distribution with  $D(i, j) \sim \exp[-(i-i_0)^2/\sigma_1 - (j-j_0)^2/\sigma_2]$ . Here,  $\sigma_1$  and  $\sigma_2$  control the width of the distribution and  $(i_0, j_0)$  is the coordinate of the lattice center, i.e.,  $(i_0, j_0) = (N_i/2, N_i'/2)$ . This non-homogeneous distribution of weights between layers could correspond to a higher density of nearby input cells. Imposing a Gaussian distribution of synaptic weights will change the eigenvalue spectrum of the covariance matrix. As a consequence, eigenfunctions with originally smaller eigenvalues can appear earlier. Furthermore, if the Gaussian distribution is rotationally symmetric, the symmetry broken by the rectangular lattice is reinstated and cells with varying orientation can appear. However, if the Gaussian distribution is not rotationally symmetric (i.e., if  $\sigma_1/\sigma_2 \neq 1$ ), the orientation of receptive fields is again pre-determined, due to imposed symmetry axes. In this case, degeneracy of eigenvalues is broken and mixing of eigenfunctions does not occur.

Figure 5 displays contour plots of the receptive fields of the first eight output cells after 10000 learning cycles (from left to right and top to bottom). Solid lines correspond to positive, dashed lines to negative synaptic weights. The input lattice was square, with  $20 \times 20$  units. We imposed a Gaussian distribution of synaptic weights with parameters  $\sigma_1 = 12$  and  $\sigma_2 = 15$ . Learning parameters  $\eta$  and  $\mu$  were equal to 0.05 and 0.1, respectively. Due to the non symmetric Gaussian distribution of weights, all units have slightly elongated receptive fields. The first unit corresponds to



**Fig. 5.** From left to right and top to bottom: contour plots of receptive fields of output units 1–8 in the case of a square lattice of  $20 \times 20$  input units. The synaptic distribution  $D(i, j)$  between layers was Gaussian with  $\sigma_1 = 12$  and  $\sigma_2 = 15$ . Solid lines indicate positive, dashed lines negative weights. The number of learning cycles was 10000, the learning parameters  $\eta$  and  $\mu$  were equal to 0.05 and 0.1, respectively

a simple cell with all-inhibitory synaptic weights. The receptive fields of the second and third units display an excitatory and an inhibitory region and resemble simple cells, selective to edges of a fixed orientation. The fourth and sixth units have receptive fields with two zero-crossings, corresponding to simple cells, selective to bars of a fixed orientation. The seventh unit is as well orientation selective, with four alternating excitatory and inhibitory regions. This unit would maximally respond to two parallel lines or bars with fixed distance and orientation. All the described units have receptive fields that resemble recorded receptive fields of simple cells in the primary visual cortex (Jones et al. 1987a–c). Up to now, there has not been any experimental evidence for receptive fields of the type of the fifth and eighth units, displaying four and six lobes. However, if the scheme of spatial information processing in terms of a local Fourier analysis is correct, such receptive fields might exist in the visual cortex.

We asked ourselves what feature detectors would develop if we presented spatially non-homogeneous, colored patterns to the network. We have found that, if such patterns are presented, color opponent feature detectors develop which combine the characteristics depicted in Figs. 2 and 5. For example, red-green and blue-yellow edge and bar detectors appear beside achromatic feature analysers.

## 5 Conclusions

We have presented a self-organizing scheme for a two-layered network of linear neurons. An anti-Hebbian rule for lateral, hierarchically organized connections between the output units renders cells with uncorrelated activities. Therefore, following a Hebbian rule, the weights between layers converge so as to form detectors of mutually independent features contained

in a presented pattern. The learning scheme is equivalent to a principal component analysis, yielding all principal components of the covariance matrix of presented patterns. Therefore, the proposed model yields a powerful mechanism for transmitting information in a most efficient way. We have tested our model with physiologically motivated input patterns. Presenting a set of color stimuli to the network leads to the formation of one achromatic and two color-opponent feature detectors with response properties similar to the results of psychophysical experiments. When the net organizes itself upon presentation of spatially varying patterns, the receptive fields of the emerging output units resemble receptive fields of simple cells found in the striate cortex.

*Acknowledgements.* One of the authors (JR) wishes to thank Art Carlson and Paul Tavan for many helpful discussions and critical comments. This work has been supported by the Bundesministerium für Forschung und Technologie (ITR-8800-G9).

*Note Added.* After completion of this work, a similar model has been proposed (Földiák 1989), characterized by symmetric lateral connections. As opposed to our model, the weight vectors do not converge to the eigenvectors of the covariance matrix. We have checked the case of symmetric connections; we were unable to find parameter sets, for which the simulations led to completely uncorrelated cells.

## References

- Ahnel PK, Kolb H, Pflug R (1987) Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina. *J Comp Neurol* 255:18–34
- Baldi P, Hornik K (1989) Neural networks and principal component analysis: learning from examples without local minima. *Neural Net* 2:53–58
- Buchsbaum G, Gottschalk A (1983) Trichromacy, opponent colours coding and optimum information transmission in the retina. *Proc R Soc London B* 220:89–113
- Campbell FW, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. *J Physiol (London)* 197:551–566
- Cowan WM (1979) The development of the brain. *Sci Am* 241:107–117
- Daugman JG (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J Opt Soc Am A* 2:1160–1169
- De Valois RL, Abramov I, Jacobs GH (1966) Analysis of response patterns of LGN cells. *J Opt Soc Am* 56:996–977
- Földiák P (1989) Adaptive network for optimal linear feature extraction. Proceedings of the IEEE International Conference on Neural Networks, Washington D.C., June 1989 I:401–405
- Hebb DO (1949) The organization of behavior. Wiley, New York
- Hopfield JJ, Feinstein DI, Palmer RG (1983) “Unlearning” has a stabilizing effect in collective memories. *Nature* 304:158–159
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J Physiol* 160:106–154

- Hubel DH, Wiesel TN (1979) Brain mechanisms of vision. *Sci Am* 241:130-144
- Jones JP, Palmer LA (1987a) The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 58:1187-1211
- Jones JP, Stepnoski A, Palmer LA (1987b) The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 58:1112-1232
- Jones JP, Palmer LA (1987c) An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* 58:1133-1258
- Judd DB (1949) Response functions for types of vision according to the Müller theory. *J Res Nat Bur Standards (Washington DC)* 42:1-3
- Kohonen T (1982) Self-organization and associative memory. Springer, Berlin Heidelberg New York
- Lawley DN, Maxwell AE (1963) Factor analysis as a statistical method. Butterworths, London
- Linsker R (1988) Self-organization in a perceptual network. *IEEE Computer*, March (1988) 105-117
- Livingstone MS, Hubel DH (1984) Anatomy and physiology of a color system in the primate visual cortex. *J Neurosci* 4:309-356
- MacKay DM (1981) Strife over visual cortical functions. *Nature* 289:117-118
- Marcelja S (1980) Mathematical description of the responses of simple cortical cells. *J Opt Soc Am* 70:1297-1300
- Oja E (1982) A simplified neuron model as a principal component analyzer. *J Math Biol* 15:267-272
- Pollen DA, Lee JR, Taylor JH (1971) How does the striate cortex begin the reconstruction of the visual world? *Science* 173:74-77
- Rubner J, Tavan P (1989) A self-organizing network for principal component analysis. *Europhys Lett* (in press)
- Rumelhart DE, Zipser D (1985) Feature discovery by competitive learning. *Cogn Sci* 9:75-112
- Schnapf JL, Kraft TW, Baylor DA (1987) Spectral sensitivity of human cone photoreceptors. *Nature* 325:439-441
- Von der Malsburg C (1973) Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* 14:85-100
- Wyszecki G, Stiles WS (1982) Color science. Wiley, New York
- Yuille AL, Kammen DM, Cohen DS (1989) Quadrature and the development of orientation selective cortical cells by Hebb rules. *Biol Cybern* 61:183-194

Received: September 26, 1989

Accepted: October 6, 1989

Dr. Jeanne Rubner  
 Physik Department  
 Theoretische Physik  
 TH München  
 James-Franck-Strasse  
 D-8046 Garching  
 Federal Republic of Germany