

Learning the Nonlinearity of Neurons from Natural Visual Stimuli

Christoph Kayser

kayser@ini.phys.ethz.ch

Konrad P. Körding

koerding@ini.phys.ethz.ch

Peter König

peterk@ini.phys.ethz.ch

Institute of Neuroinformatics, University / ETH Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Learning in neural networks is usually applied to parameters related to linear kernels and keeps the nonlinearity of the model fixed. Thus, for successful models, properties and parameters of the nonlinearity have to be specified using a priori knowledge, which often is missing. Here, we investigate adapting the nonlinearity simultaneously with the linear kernel. We use natural visual stimuli for training a simple model of the visual system. Many of the neurons converge to an energy detector matching existing models of complex cells. The overall distribution of the parameter describing the nonlinearity well matches recent physiological results. Controls with randomly shuffled natural stimuli and pink noise demonstrate that the match of simulation and experimental results depends on the higher-order statistical properties of natural stimuli.

1 Introduction

One important application of artificial neural networks is to model real cortical systems. For example, the receptive fields of sensory cells have been studied using such models, most prominently in the visual system. An important reason for the choice of this system is our good knowledge of natural visual stimuli (Ruderman, 1994; Dong & Atick, 1995; Kayser, Einhäuser, & König, in press). A number of recent studies train networks with natural images, modeling cells in the primary visual cortex (see Simoncelli & Olshausen, 2001, and references therein). In these studies, a nonlinear model for a cell was fixed, and the linear kernel was optimized to maximize a given objective function. The resulting neurons share many properties of either simple or complex cells as found in the visual system. However, these studies incorporate prior knowledge about the neuron model to quantify the nonlinearity of the neuron in advance. But to better understand and predict properties of less-well-studied systems, where such prior knowledge is

not available, models are needed that can adapt the nonlinearity of the cell model.

Here we make a step in this direction and present a network of visual neurons in which both a parameter controlling the nonlinearity and the linear receptive field kernels are learned simultaneously. The network optimizes a stability objective (Földiak, 1991; Kayser, Einhäuser, Dümmer, König, & Körding, 2001; Wiskott & Sejnowski, 2002; Einhäuser, Kayser, König, & Körding, 2002) on natural movies, and the transfer functions learned are comparable to those found in physiology as well as those used in other established models for visual neurons.

2 Methods

As a neuron model, we consider a generalization of the two subunit energy detector (Adelson & Bergen, 1985). The activity A for a given stimulus I is given by

$$A(t) = (|I \bullet W_1|^N + |I \bullet W_2|^N)^{\frac{1}{N}}. \quad (2.1)$$

Each neuron is characterized by two linear filters $W_{1,2}$, as well as the exponent of the transfer function N . For $N = 1$, the transfer function has unit gain, and except for the absolute value, the neuron performs a linear operation. For $N = 2$, this corresponds to a classical energy detector. All of these parameters are subject to an optimization of an objective function. The objective used here is temporal coherence, a measure of the stability of the neuron's output plus a decorrelation term:

$$\begin{aligned} \Psi = \Psi_{time} + \Psi_{Decor} = & - \frac{\langle (A(t) - A(t - \Delta t))^2 \rangle_{\text{Stimuli}}}{\text{var}(A(t))_{\text{Stimuli}}} \\ & + \sum_{\text{Neurons } i,j} -\langle A_i A_j \rangle_{\text{Stimuli}}^2. \end{aligned} \quad (2.2)$$

A neuron is optimal with respect to the first term of the objective function if its activity changes slowly on a timescale given by Δt . The second term avoids a trivial solution with identical receptive fields of all neurons. Implementing optimization by gradient ascent, the formula for the change of parameters can be obtained by differentiation of equation 2.2 with respect to the parameters describing the neurons.

The subunits' receptive fields W are initialized with values varying uniformly between 0 and 1 and the exponents with values between 0.1 and 6. During optimization, in a few cases, the exponent of a neuron would diverge to infinity. In this case, the transfer function implements a maximum operation on the subunits, and the precise value of the exponent is no longer relevant. To avoid the associated numerical problems, we constrain the exponent to be smaller than 15.

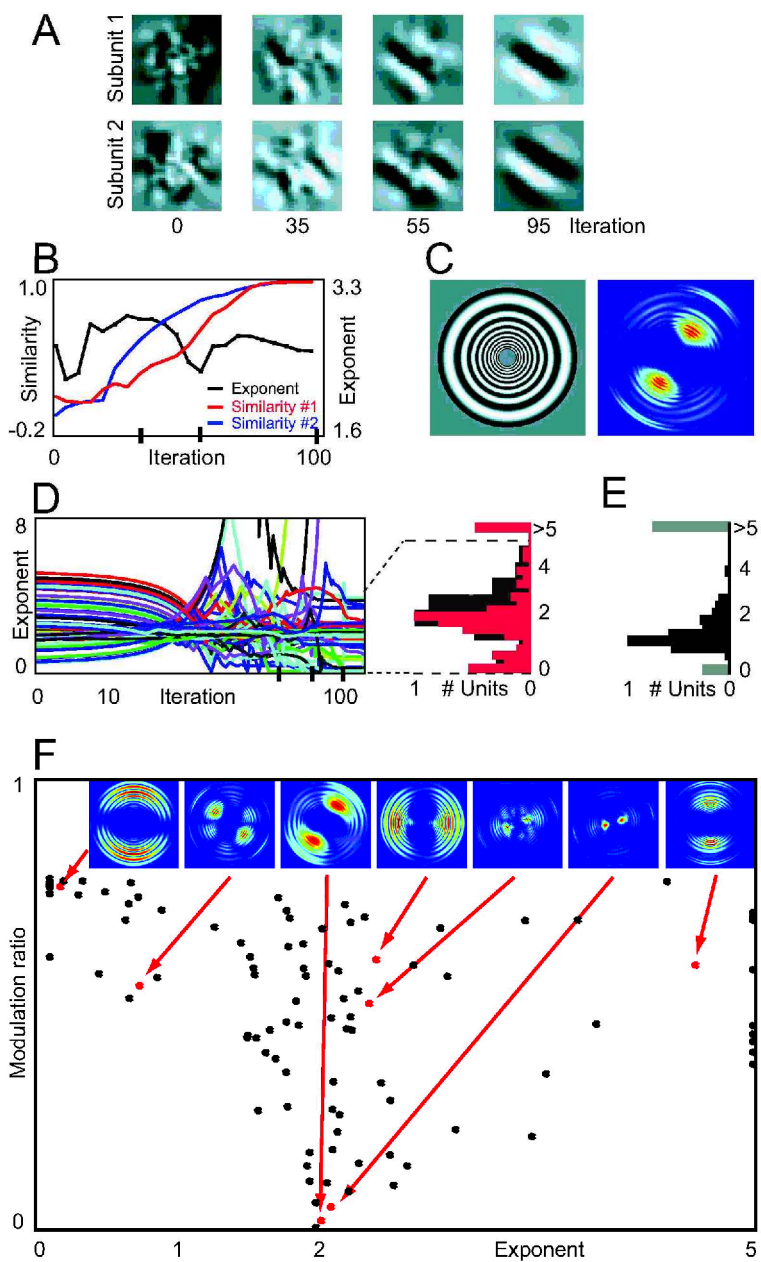
The input stimuli are taken from natural movies recorded by a light-weight camera mounted to the head of a cat exploring the local park (for details, see Einhäuser et al., 2002). As controls, two further sets of stimuli are used. First, temporally “white” stimuli are constructed by randomly shuffling the frames of the natural video. Second, the natural movie is transformed into spatiotemporal “pink” noise by assigning random phases to the space-time Fourier transform. This yields a movie with the same second-order statistics as the original but without higher-order correlations. From two consecutive video frames, one corresponding to $t - \Delta t$, the other to t , we cut pairs of patches of size 30×30 pixels (≈ 6 degrees). The square patches are multiplied by a circular gaussian window in order to ensure isotropic input. The training set consists of 40,000 such pairs. The dimension of stimulus space is reduced by a principal component analysis in order to ease the computational load. Excluding the first component corresponding to the mean intensity, we keep the first 120 components out of a total of 1800, carrying more than 95% of the variance.

We quantify the properties of the receptive fields in an efficient way by convolving the subunits with a large circular grating covering frequencies ranging from 0.5 cycles per patch to 15 cycles per patch and all orientations (see Figure 1C). The convolutions of the subunits with this patch are summed pointwise according to the cell model (see equation 2.1) resulting in a two-dimensional activity diagram.

3 Results

We optimize the receptive fields and exponents of a population of 80 cells. In the following, we first discuss one example in detail before reporting results of the whole population.

The development of the subunits of the example neuron is shown in Figure 1A. It is selective to orientation and spatial frequency. This selectivity develops after about 50 iterations simultaneously in both subunits, and the spatial profile is very similar to a Gabor filter. Indeed the best-fitted Gabor filter explains 81% and 74% of the variance of the two subunits’ receptive fields. Simultaneously with their emergence, the oriented structures in the two subunits are shifted relative to each other by 87 degrees. The evolution of the exponent of this neuron during optimization is shown in Figure 1B. It converges to a value of 2.35, which is close to the exponent of 2.0 of the classical energy detector. The response profile characterizing the spatial receptive field (see methods) is displayed in Figure 1C. It shows that the neuron is selective to orientation (oblique) and spatial frequency (11 pixel wavelength). The response to a grating shows a constant increase toward higher spatial frequencies, with a small-frequency doubled modulation on top. The response to a drifting grating (data not shown) shows a similar effect, which is caused by the absolute value in the formula for the activity. However, it is weak compared to the constant increase of the response.



The neuron's small modulation ratio for drifting gratings ($F1/F0$: 0.26) is equivalent to a translation-invariant response. Therefore, this neuron shares the properties of complex cells found in primary visual cortex (Skottun et al., 1991).

Looking at the spatial tuning properties of the cells in the entire population, we find that all neurons are selective to orientation (mean width at half height 35 degrees and the ratio of the response strength at preferred orientation to the response at orthogonal orientation is 11.2). Furthermore, the neurons are selective to spatial frequency with a median spatial frequency selectivity index of 59 (for a definition, see Schiller, Finlay, & Volman, 1976).

In the population, independent of the initial value, most exponents converge quickly toward a value close to two. Then, for most cells during the next iterations, the exponent changes only little. Sometimes, however, the exponent suddenly explodes and grows until it reaches our constraint of $N = 15$. Over the whole population, most cells acquire an exponent close to 2 (see Figure 1D right, red), although the distribution is fairly broad. This shows that the classical energy detector is an optimal solution for the given objective. Furthermore, in comparison to recent physiological results obtained in cat primary visual cortex, the distribution of exponents is similar (see Figure 1D right, gray, reproduced from Lau, Stanley, & Dan, 2002). An overview of the relation between the exponent of a neuron and its response properties for the whole population is given in Figure 1F. There, we plot the exponent versus the modulation ratio for drifting gratings. Note that cells with a low modulation ratio would be classified as complex cells. Responses of cells with a high modulation ratio are specific also to the position of a stimulus, as are simple cells in visual cortex. Cells that can be classified as complex cells are found only in the region of the exponent close to two. Cells

Figure 1: *Facing page*. (A) Subunits' receptive fields of an example neuron as a function of the number of iterations. (B) The exponent N and the similarity index of the subunits for the example neuron during the optimization. The similarity index is the scalar product of the final subunit (after 100 iterations) with the subunit at a given iteration divided by the lengths of these vectors. The tick marks on the x -axis indicate 35, 55, and 95 iterations in correspondence to panel A. (C) Left: The circular grating patch used for quantifying the neurons response properties. Right: Response diagram of the example neuron on a color scale from bright red (high activity) to dark blue (no activity). (D) Left: The evolution of the exponents for the population. The tick marks on the x -axis indicate 35, 55, and 95 iterations in correspondence to panel A. Right: The histogram of the exponents after 100 iterations over the population is shown in red. Physiological data taken from Lau et al. (2002) are shown in gray. (E) Histograms of the final values for the exponents in the control simulations. Light gray: temporally white stimulus; black: spatiotemporal pink noise. (F) Exponent vs. modulation ratio for drifting gratings. The response diagrams of seven neurons are shown at the upper border. On the right border, all neurons are located with an exponent larger than 5.

with large modulation ratios, on the other hand, are found for all values of the exponent. Furthermore, the preferred orientation and spatial frequency are distributed homogeneously.

To probe which properties of natural scenes are important for the stable development of the exponent and the spatial receptive fields, we perform two controls. Since the objective function is based on temporal properties of the neurons' activity, we first use a stimulus without temporal correlations. This temporal white stimulus is constructed by shuffling the frames of the original video (see section 2), but preserves its spatial structure. After a few iterations, the exponents of all neurons diverge to either the upper or lower limit imposed (see Figure 1E, gray). The spatial receptive fields appear noisy and display weak tuning to orientation (median of the ratio of response at optimal orientation to response at orthogonal orientation is 3.7). This shows that the distribution of exponents obtained in the first simulation is not a property inherent in our network. Rather, a smooth temporal structure of the input is necessary for the development of both the exponent and the spatial receptive fields.

As a second control, we train the network on spatiotemporal pink noise (see section 2) having the same second-order statistics as the natural movie. While this stimulus lacks the higher-order structure of natural scenes, it has a smooth, temporal structure compared to the first control. During optimization, the exponents converge as in the original simulation. However, the histogram of final values (see Figure 1E, black) differs markedly from that in Figure 1D. Most neurons acquire an exponent slightly above one, with many neurons having a linear slope of their transfer function. Furthermore, most neurons resemble low-pass filters, and their orientation tuning is weak (median of the response at preferred orientation to response at the orthogonal orientation is 2.9). Thus, the second-order correlations of natural movies are sufficient to allow stable learning of the exponent, as well as organized spatial receptive fields. However, to achieve a good match to physiological results in terms of both spatial receptive fields and the exponent of the transfer function, the higher-order structure of the natural scenes is decisive.

4 Discussion

We report here a network simultaneously optimizing the linear kernel of the given neuron model and the exponent of the neurons' transfer functions with respect to the same objective function.

The results of the study are interesting for the following reasons. First, by changing parameters controlling the transfer function, neurons can represent a much larger class of nonlinear functions and, given a fixed network size, are able to solve a larger class of problems. Second, nonlinearities are ubiquitous in sensory systems (Ghazanfar, Krupa, & Nicolelis, 2001; Shimegi, Ichikawa, Akasaki, & Sato, 1999; Anzai, Ohzawa, & Freeman, 1999;

Lau et al., 2002; Escabi & Schreiner, 2002; Field & Rieke, 2002) and a wide variety of effects contributes to nonlinear effects within a neuron (Softky & Koch, 1993; Mel, 1994; Reyes, 2002). Yet in most cases, we simply lack the knowledge for a quantitative specification. Therefore, to model these systems, networks are required that can adjust all parameters and do not require detailed a priori knowledge.

While most learning schemes still concentrate on linear kernels, a small number of studies address optimizing the neurons nonlinearity. Bell and Sejnowski (1995) start from an Infomax principle and derive a two-step learning scheme, which allows optimizing the shape of a (sigmoidal) transfer function to match the input probability density. This directly relates to the concept of independent component analysis, where the adaptation of the transfer function adjusts the nonlinearity of the system (Nadal & Parga, 1994; Everson & Roberts, 1998; Obradovic & Deco, 1998; Pearlmutter & Parra, 1997). Recently, this approach was applied to natural auditory stimuli comparing the resulting receptive field structure to physiological results (Lewicki, 2002). However, some of these theoretical derivations require an invertible transfer function, which is not the case in our network and might not apply in general for neurons that discard information, such as neurons having invariance properties. Extending this previous research, we optimize the weight matrix and the nonlinearity simultaneously at each iteration and compare the resulting distribution of the parameters to recent experimental data.

In our network, we find many cells that are selective to orientation and spatial frequency, have an exponent around 2, and have translation-invariant responses. Their properties match the classical energy detector model for complex cells (Movshon, Thompson, & Tolhurst, 1978; Adelson & Bergen, 1985). Thus, our results show how this classical model of complex neurons can develop in a generalized nonlinear model. Furthermore, in a recent study of the cat visual system, the nonlinearity of the computational properties of cortical neurons has been investigated (Lau et al., 2002). The measured exponent of the nonlinear transfer function shows a large variation. An exponent around 2 best explains the response of many neurons, while in other cells it ranges between 0 and high values. The overall distribution of exponents found in the study examined here matches these experimental results surprisingly well.

Acknowledgments

We thank Laurenz Wiskott for introducing the evaluation of receptive fields using the circular grating patch to us, and Mathew Diamond and Rasmus Petersen for valuable discussion of yet unpublished results. This work was supported by the Centre of Neuroscience Zürich (C.K.), the SNF (Grant No. 31-65415.01, P.K.) and the EU IST-2000-28127/BBW 01.0208-1 (K.P.K., P.K.).

References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Optic. Soc. Am. A*, *2*, 284–299.
- Anzai, A. I., Ohzawa, I., & Freeman, R. (1999). Neural mechanisms for processing binocular information I. Simple cells. *J. Neurophysiol.*, *82*, 891–908.
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1004–1034.
- Dong, D. W., & Atick, J. J. (1995). Statistics of natural time varying images. *Network: Computation in Neural Systems*, *6*, 345–358.
- Einhäuser, W., Kayser, C., König, P., & Körding, K. P. (2002). Learning the invariance properties of complex cells from their responses to natural stimuli. *Eur. J. Neurosci.*, *15*, 475–486.
- Escabi, M. A., & Schreiner, C. E. (2002). Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain. *J. Neurosci.*, *22*, 4114–4131.
- Everson, R. M., & Roberts, S. J. (1998). Independent component analysis: A flexible nonlinearity and decorrelating manifold approach. *Neural Computation*, *11*, 1957–1983.
- Field, G. D., & Rieke, F. (2002). Nonlinear signal transfer from mouse rods to bipolar cells and implications for visual sensitivity. *Neuron*, *34*, 773–785.
- Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, *3*, 194–200.
- Ghazafar, A. A., Krupa, D. J., & Nicolelis, M. A. L. (2001). Role of cortical feedback in the receptive field structure and nonlinear response properties of somatosensory thalamic neurons. *Exp. Brain Res.*, *141*, 88–100.
- Kayser, C., Einhäuser, W., Dümmer, O., König, P., & Körding, K. P. (2001). Extracting slow subspaces from natural videos leads to complex cells. In G. Dorffner, H. Bischoff, & K. Kornik (Eds.), *Proc. Int. Conf. Artif. Neural Netw. (ICANN)* (pp. 1075–1080). Berlin: Springer-Verlag.
- Kayser, C., Einhäuser, W., & König, P. (in press). Temporal correlations of orientations in natural scenes. *Neurocomputing*.
- Lau, B., Stanley, G. B., & Dan, Y. (2002). Computational subunits of visual cortical neurons revealed by artificial neural networks. *Proc. Natl. Acad. Sci. USA*, *99*, 8974–8979.
- Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neurosci.*, *5*(4), 356–363.
- Mel, B. (1994). Information processing in dendritic trees. *Neural Computation*, *6*, 1031–1085.
- Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Receptive field organization of complex cells in the cat's striate cortex. *J. Physiol. (London)*, *283*, 79–99.
- Nadal, J.-P., & Parga, N. (1994). Nonlinear neurons in the low-noise limit: A factorial code maximized information transfer. *Network: Computation in Neural Systems*, *5*, 565–581.
- Obrovic, D. & Deco, G. (1998). Information maximization and independent component analysis: Is there a difference? *Neural Computation*, *10*(8), 2085–2101.

- Pearlmutter, B., & Parra, L. (1997). Maximum likelihood blind source separation: A context-sensitive generalization of ICA. In M. Mozer, J. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems*, 9 (pp. 613–619). Cambridge, MA: MIT Press.
- Reyes, A. (2002). Influence of dendritic conductances on the input-output properties of neurons. *Annu. Rev. Neurosci.*, 24, 653–675.
- Ruderman, D. L. (1994). The statistics of natural images. *Network: Computation in Neural Systems*, 5, 517–548.
- Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single cell properties in monkey striate cortex. III. Spatial frequency. *J. Neurophysiol.*, 39, 1334–1351.
- Shimegi, S., Ichikawa, T., Akasaki, T., & Sato, H. (1999). Temporal characteristics of response integration evoked by multiple whisker stimulations in the barrel cortex of rats. *J. Neurosci.*, 19, 10164–10175.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Ann. Rev. Neurosci.*, 24, 1193–1216.
- Skottun, B. C., De Valois, R. L., Grosf, D. H., Movshon, J. A., Albrecht, D. G., & Bonds, A. B. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision Res.*, 31, 1079–1086.
- Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.*, 13, 334–350.
- Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, 14, 715–770.

Received September 4, 2002; accepted January 30, 2003.