Exploring the Role of Intrinsic Plasticity for the Learning of Sensory Representations

Nicholas J. Butko[1] and Jochen Triesch[1,2]

1- University of California, San Diego - Cognitive Science
9500 Gilman Dr. 0515, La Jolla, CA, 92093-0515 - USA

2- Frankfurt Institute for Advanced Study
Max-von-Laue-Str. 1, 60438 Frankfurt am Main - Germany

Abstract.   Intrinsic plasticity (IP) refers to a neuron's ability to regulate its firing activity by adapting its intrinsic excitability. Previously, we showed that model neurons combining IP with Hebbian synaptic plasticity can adapt their weight vector to discover heavy-tailed directions in the input space. In this paper we consider networks of coupled model neurons and show how a population of such units can solve a standard non-linear ICA problem. We also present a simple model for the formation of maps of oriented receptive fields in primary visual cortex. Together, our results indicate that intrinsic plasticity may play an important role for learning efficient representations in populations of cortical neurons.

1   Introduction

Biological neurons regulate their firing activity by adapting their intrinsic excitability. Such *intrinsic plasticity* (IP) seems to be a ubiquitous phenomenon in the brain [1]. For example, Desai *et al.* showed that neurons that had been prevented from spiking for two days increased their response to current injection [2]. It is frequently assumed that IP contributes to the homeostasis of a neuron's firing activity. Baddeley *et al.* found that neurons in visual cortical areas show exponential distributions of their firing rate, which is thought to maximize a neuron's information transfer given a fixed energy budget [3]. This is because the exponential distribution has the maximum entropy among all distributions of a positive random variable with a fixed mean. It has been speculated that IP may be instrumental in achieving approximately exponential firing rate distributions in cortical neurons [4]. We have recently shown that IP that drives a neuron to exhibit an exponential firing rate distribution can synergistically interact with Hebbian learning at the synapses. These two processes lead to the discovery of heavy-tailed directions in the input space [5]. In this paper we extend these results to populations of neurons with IP. Our specific goal is to explore the potential role of IP for learning efficient map-like representations for sensory stimuli.

Computational models of the emergence of sensory representations in the brain abound. Frequently, they fall into one of two categories: *functional models* or *mechanistic models*. Mechanistic models start from neuroscientific data about the structure of cortical networks and cortical plasticity mechanisms (cell types,

connection patterns, plasticity rules, ...) which are distilled into simplified models. These models are trained on actual sensory data or noise patterns and the learned representations can be compared to neuroscientific data. If the resulting representations are similar to those found in the brain then this provides evidence that the processes in the brain have been accurately captured, but it does not clarify why the brain operates this way or in what sense the brain's solution may be optimal. An example of a model of this kind by Linsker is [6], where V1-style orientation columns are learned from random prenatal visual noise through Hebbian learning. Later Miller extended this work to learn many of the various map-structures in V1, and used model neurons that were somewhat more plausible [7].

Functional models focus on the abstract computational goal of the problem. For the case of learning sensory representations they start by asking: what is the *optimal* way to represent sensory stimuli, e.g. natural images, where optimality is usually defined with respect to certain statistical criteria (e.g., sparsness, independence, temporal coherence, ...) and additional constraints. Algorithms are derived to learn the optimal solution to the problem which can again be compared to neuroscientific data. If the found solution resembles the biological solution, then this provides evidence that the brain may in fact be trying to optimize a similar objective function. Through what mechanisms the brain may achieve this goal is typically not answered, however.

A central idea in many functional models is information maximization. According to this idea, individual neurons should spread out their responses in dense regions of the input space and compress responses in sparse regions. In effect, this maps the input to a uniform output distribution, maximizing entropy. Laughlin showed that blowfly Large Monopolar Cells have been adapted so that their input/output transfer functions nearly optimally represent the contrast statistics of their environment [8]. Bell & Sejnowski showed that the same information maximization principle can be applied to the independent component analysis (ICA) problem. They applied their technique to natural images and found oriented, bandpass sources [9] similar to those observed in V1. Olshausen & Field showed that oriented, bandpass receptive fields also arise when optimizing image reconstruction error subject to lifetime sparseness constraints [10]. They imposed a sparse prior on the contribution of each basis function in a generative model with the intuition that among the space of possible sources of an image, each one is present only rarely. This intuition is confirmed by Ruderman, who showed that bandpass filter responses to natural images follow an exponential distribution [11].

The model we present in the following attempts to bridge the gap between mechanistic and functional models. On the one hand, it has a clear connection to the idea of information maximization [5]. On the other hand, it has a mechanistic formulation that is biologically plausible because it makes use of information that is local in time and space, and uses patterns of lateral connections characteristic of neural populations. While similar bridges have been attempted before, e.g. [12], our model is the first to utilize IP as a fundamental

mechanism for the learning of sensory representations.

2 Network Model with Intrinsic Plasticity

We consider a network of units learning to represent a sensory input signal $\mathbf{x}$. The activity of unit i in the network is given by:

$$y_i(h_i) = [1 + \exp(-a_i h_i - b_i)]^{-1}, \text{ where } h_i = \mathbf{x} \cdot \mathbf{w}_i , \qquad (1)$$

where $\mathbf{w}_i$ is the neuron's weight vector, and a_i and b_i are adjustable parameters of the neuron's transfer function that are controlled by IP. In particular a_i and b_i are subject to the following learning rule:

$$\Delta a_i \overset{(a)}{=} \eta_{\text{IP}} \left[a_i^{-1} + h_i - (2 + \mu^{-1}) h_i y_i + \mu^{-1} h_i y_i^2 \right]$$
$$\Delta b_i \overset{(b)}{=} \eta_{\text{IP}} \left[1 - (2 + \mu^{-1}) y_i + \mu^{-1} y_i^2 \right] , \qquad (2)$$

where η_{IP} is a small learning rate and μ is the desired mean activity of all units. As derived in [13], this learning rule has the effect of making the distribution of y_i a sparse, approximately exponential distribution, thereby maximizing the unit's entropy given a fixed average activity. Note that this rule is *local* in space and time, making it physiologically plausible.

Plasticity of the weight vectors $\mathbf{w}_i$ is modeled with a Hebbian learning rule. In [5], we considered a single unit learning rule of the form $\Delta \mathbf{w} \propto \mathbf{x} y$. We showed that the coupling of IP with this form of Hebbian learning allowed the unit to discover heavy-tailed directions in the input. To extend this model to a population of model neurons, we introduce a *neighborhood function* $\mathcal{N}$ as commonly used in self-organizing maps. The value of the neighborhood function for neuron i is determined by its activity y_i and the activities of all other neurons, i.e. $\mathcal{N}(y_i; \mathbf{y})$. Specific forms of $\mathcal{N}$ are introduced below. After each stimulus presentation, the weights are updated according to:

$$\Delta \mathbf{w}_i \overset{(a)}{=} \mathbf{x} y \mathcal{N}(y_i; \mathbf{y}), \quad \mathbf{w}_i \overset{(b)}{\Leftarrow} \frac{\mathbf{w}_i + \eta_{\text{Hebb}} \Delta \mathbf{w}_i}{\| \mathbf{w}_i + \eta_{\text{Hebb}} \Delta \mathbf{w}_i \|} . \qquad (3)$$

where η_{Hebb} is a learning rate, $\Leftarrow$ denotes assignment, and the normalization in (3b) mimics competition between synapses on a neuron's dendritic tree [7].

3 Experiments

3.1 The "bars" problem

The bars problem is standard non-linear ICA problem introduced by Földiák [14]. Horizontal and vertical bars are presented on an R-by-R retina. The presence or absence of a bar is independent of that of any other bars. The unsupervised learning problem is to learn filters that correspond to the individual independent components, *i.e.* the bars. The problem is non-linear because the

Fig. 1: A population of 20 model neurons has learned all independent sources in the bars problem. The weight vector of each unit has discovered a single bar. Parameters were: $\beta = 0.2$, $\eta_{\text{Hebb}} = 0.01$, $\eta_{\text{IP}} = 0.005$, and $\mu = 0.1$.

pixel at the intersection of two bars is just as bright as any other pixel of the bars, not twice as bright. In our previous work [5], we showed that a single model neuron with IP and Hebbian learning discovers one of the bars when exposed to stimuli from the bars problem. Here we use a population of units to learn the complete problem. We use a retina of 10-by-10 pixels and the probability of any of the 20 bars occuring in a given stimulus is 10%. Since we want filters that respond highly when bars are present and not otherwise, the desired mean firing rate μ is set at 10%. $\mathcal{N}$ is chosen to enforce a winner-take-all competition between the units, so that the maximally activated neuron updates its weight vector in a standard Hebbian fashion, and all other units update their weight in an anti-Hebbian manner regulated by a decorrelation parameter β:

$$\mathcal{N}_{\text{bars}}(y_i; \mathbf{y}) = \left\{ \begin{array}{l} 1 : y_i = \max(\mathbf{y}) \\ -\beta : \text{else} \end{array} \right. . \qquad (4)$$

All units update their intrinsic parameters independently, as described in Equation 2(a&b). Figure 1 shows the parameters used and the results of learning. With the specified parameters, we found that learning was completely successful in 19 out of 20 trials. Varying the learning rates η_{Hebb} and η_{IP} affected learning little, provided both remained above 0. When μ was 0.05, redundant filters were learned, and when it was 0.2, multiple bars were represented within single filters. This suggests that when the true mean of the components is unknown, it may be a better strategy to choose μ too high rather than too low. This way, all true sources will likely be captured because individual filters each learn to represent several sources. Learning substantially worsened when β was less than 0.05 or greater than 0.3, with effects similar to μ too low or too high respectively.

3.2   Modeling the Emergence of Orientation Maps

Receptive fields of simple cells in V1 are oriented, localized, and bandpass. Neurons are arranged in a columnar fashion that reflects their orientation preference. For modelling the emergence of orientation columns, we consider our populations of neurons to be located on a two-dimensional sheet, with neuron i at grid position $(j, k)_i \in \mathbb{N} \times \mathbb{N}$ after the fashion of a Kohonen feature map. The most active unit exhibits a center-surround influence on learning in its neighbors according to a difference of Gaussians (DoG) neighborhood function centered at
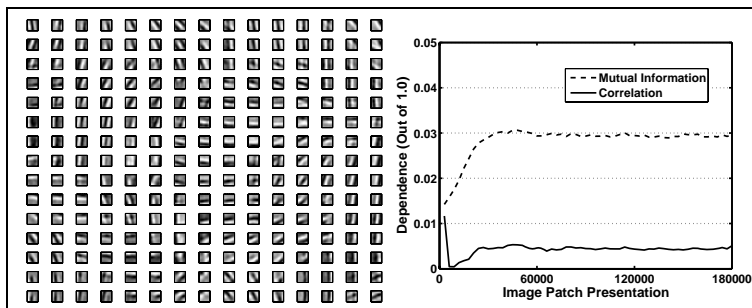
Fig. 2: **Left:** Receptive fields learned from natural image patches. Parameters were: $\eta_{\mathrm{Hebb}} = 0.05$, $\eta = 0.01$, $\mu = 0.15$, $\sigma_c = 1$, $\sigma_s = 1.5$. **Right:** Mean correlation and mean normalized mutual information $(2 \cdot \mathcal{I}(X;Y)/[\mathcal{H}(X)+\mathcal{H}(Y)]$, 6 bins) between firing activity of each unit pair remained very low.

the coordinates of the most activated neuron (j_*, k_*):

$$
\mathcal{N}_{\mathrm{map}}(y_i; \mathbf{y}) \;=\; \frac{1}{2\pi\sigma_c^2} \exp\left(\frac{(j_i - j_*)^2 + (k_i - k_*)^2}{2\sigma_c^2} \right)
$$
$$
- \frac{1}{2\pi\sigma_s^2} \exp\left(\frac{(j_i - j_*)^2 + (k_i - k_*)^2}{2\sigma_s^2} \right) \;, \qquad (5)
$$

where σ_c and σ_s determine the range of the center and surround interaction.

Our inputs were taken from natural images collected by Van Hateren [15]. We used log-intensity images because these have greater contrast and this transform is performed in the early visual pathway [15]. We convolved the images with a difference of Gaussians (DoG) filter to model the center-surround opponency of neurons in the lateral geniculate nucleus (LGN) [7]. 500 10x10 image patches were drawn at random from each of 375 images, and were presented once to each neuron in our population. The input had positive and negative values simulating populations of ON and OFF cells in the LGN [6].

We used a population of 225 units arranged on a 15x15 sheet. Each unit had a 10x10 receptive field size, making the population 2.25 times overcomplete. The results of learning are shown in Figure 2. Learning was robust to changes in the parameters over a wide range of values. The learned filters exhibit a variety of orientations, frequencies, and locations. Moreover, they exhibit smooth interpolation in local regions. This is reminiscent of the orientation-map structure in V1. We measured the degree of dependence among firing activities and found that on average a unit gave only 3% of the maximum possible information about any other unit in the 2.25 times overcomplete population.

4 Concluding Remarks

Different forms of plasticity are involved in shaping sensory representations in the brain and it is important to understand how these different mechanisms

interact. In [5] we developed model neurons that maintain sparse lifetime distributions through IP while performing Hebbian learning at the synapses. Here we constructed networks of such neurons whose learning was coupled using a winner-take-all mechanism or a self-organizing map architecture. In the former case, we solved a standard non-linear ICA problem, and in the latter we found maps of Gabor-like receptive fields as seen in primary visual cortex. Pairwise independence between neurons was well-maintained on average. Our results suggest that IP may play an important role in the unsupervised learning of independent, information maximizing representations in the cortex.

Acknowledgments

References

[1] W. Zhang and D. J. Linden, The other side of the engram: experience-dependent changes in neuronal intrinsic excitability. *Nature Reviews Neuroscience*, 4:885-900, 2003.

[2] N. S. Desai, L. C. Rutherford and G. Turrigiano, Plasticity in the intrinsic excitability of cortical pyramidal neurons, *Nature Neuroscience*, 2(6):515:520, 1999.

[3] R. Baddeley, L. F. Abbott, M. C. Booth, F. Sengpiel, T. Freeman, E. A. Wakeman, and E. T. Rolls, Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Roy. Soc. of Lond. B*, 264:1775-1783, 1998.

[4] M. Stemmler and C. Koch, How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nature Neuroscience*, 2:521-527, 1999.

[5] J. Triesch, Synergies between Intrinsic and Synaptic Plasticity in Individual Neurons, *Advances in Neural Information Processing Systems*, 17:1417-1424, 2005.

[6] R. Linsker, From basic network principles to neural architecture: emergence of oriented columns. *Proceedings of the National Academies of Science*, 83:8779-8783, 1986.

[7] K. D. Miller, A Model for the Development of Simple Cell Receptive Fields and the Ordered Arrangement of Orientation Columns through Activity-dependent Competition between ON- and OFF-Center Inputs, *The Journal of Neuroscience*, 14(1):409-441, 1994.

[8] S. Laughlin, A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch* 36:910-912, 1981.

[9] A. J. Bell and T. J. Sejnowski, The "Independent Components" of Scenes are Edge Filters. *Vision Research*, 37(23):3327-3338, 1997.

[10] B. A. Olshausen and D. J. Field, Sparse Coding with an Overcomplete Basis Set: A strategy employed by V1? *Vision Research*, 37(23):3311-3325, 1997.

[11] D. L. Ruderman and W. Bialek, Statistics of natural images: scaling in the woods, *Physical Review Letters* 73, 1994.

[12] L. N. Cooper, N. Intrator, B. S. Blais, and H. Z Shouval. *Theory of Cortical Plasticity*, World Scientific, London, 2004.

[13] J. Triesch, A Gradient Rule for the Plasticity of a Neuron's Intrinsic Excitability, *Proc. Int. Conf. on Artificial Neural Networks* 65-70, 2005.

[14] P. Földiák, Forming sparse representation by local anti-hebbian learning, *Biological Cybernetics*, 64:165-170, 1990.

[15] J. H. van Hateren and A. van der Schaaf, Independent component filters of natural images comparsed with simple cells in primary visual cortex. *Proc. Roy. Soc. of Lond. B*, 265:359-366, 1998.