

## Synergies Between Intrinsic and Synaptic Plasticity Mechanisms

Jochen Triesch

*triesch@fias.uni-frankfurt.de*

*Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany, and Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093-0515, U.S.A.*

**We propose a model of intrinsic plasticity for a continuous activation model neuron based on information theory. We then show how intrinsic and synaptic plasticity mechanisms interact and allow the neuron to discover heavy-tailed directions in the input. We also demonstrate that intrinsic plasticity may be an alternative explanation for the sliding threshold postulated in the BCM theory of synaptic plasticity. We present a theoretical analysis of the interaction of intrinsic plasticity with different Hebbian learning rules for the case of clustered inputs. Finally, we perform experiments on the “bars” problem, a popular nonlinear independent component analysis problem.**

### 1 Introduction ---

Synaptic learning mechanisms have always been a central area of research in neural computation, but there has been relatively little work on intrinsic plasticity (IP) mechanisms, which change a neuron’s intrinsic excitability. Such changes have been observed in many species and brain areas (Desai, Rutherford, & Turrigiano, 1999; Zhang & Linden, 2003; Daoudal & Debanne, 2003; Zhang, Hung, Zhu, Xie, & Wang, 2004; Cudmore & Turrigiano, 2004; van Welie, van Hooft, & Wadman, 2004, 2006), and they can be measured as changes to a neuron’s frequency-current (f-I) function, such as alterations to the spike threshold or the initial slope of the f-I function. Interestingly, such effects seem to occur on different timescales, with some changes occurring over days (e.g., Desai et al., 1999) and others occurring much faster (Cudmore & Turrigiano, 2004; van Welie et al., 2004, 2006). Despite recent progress, the biophysical implementation of these mechanisms remains somewhat unclear, and the exact computational functions of these forms of plasticity are not well established either. Changes to the intrinsic excitability of neurons may help to shape the dynamics of neural circuits (Marder, Abbott, Turrigiano, Liu, & Golowasch, 1996) in desired ways. It is also reasonable to assume that IP contributes to neural homeostasis by keeping the level of activity of individual neurons and neural

circuits in a desirable regime. It is less clear, however, what exactly such a desirable regime may be. It has been proposed that efficient sensory coding can be obtained by adjusting neural responses to match the statistics of signals from the environment in a way that maximizes information transmission. In particular, if a neuron uses all of its possible response levels equally often, then its entropy is maximized (Laughlin, 1981). For visual cortex of cats and macaques, Baddeley, Abbott, Booth, Sengpiel, and Freeman (1998) have observed that individual neurons exhibit an approximately exponential distribution of firing rates in response to stimulation with natural videos. They have argued that this may be motivated through the fact that the exponential distribution maximizes entropy for a fixed mean. Thus, neurons may assume exponential firing-rate distributions in order to maximize the information they convey to downstream targets given a fixed average activity level, which corresponds to a certain energy budget. Based on this idea, Stemmler and Koch (1999) developed a model of IP for a Hodgkin-Huxley-style model neuron. They derived a learning rule for the adaptation of a number of voltage-gated channels to match the firing-rate distribution of the unit to a desired distribution. Shin, Koch, and Douglas (1999) adjusted two conductances in electronic circuit analogs of neurons to adapt to the mean and variance of the time-varying somatic input current.

In this work, we describe an IP model for a simple continuous-activation model neuron with only two parameters. We then study the interaction of intrinsic and synaptic learning processes and show how a model neuron with IP at the soma and Hebbian learning at the synapses can discover heavy-tailed directions in its input. We analyze the interaction of IP with different Hebb rules including a BCM rule (Bienenstock, Cooper, & Munro, 1982). We also show how IP may function as an alternative to the sliding threshold mechanism assumed to balance long-term potentiation (LTP) and long-term depression (LTD) in BCM theory. Finally, we demonstrate how a single unit with IP and Hebbian learning discovers an independent component in the “bars” problem, a standard unsupervised learning problem.

## 2 A Model of Intrinsic Plasticity

---

Since the long-term goal of our research is to study the effects of IP at the level of large networks, we choose to use a simple rate-coding model neuron. We model IP as changes to the neuron’s nonlinear activation function. We consider a model neuron with a parametric sigmoid nonlinearity of the form

$$y = g_{ab}(x) = \frac{1}{1 + \exp(-(ax + b))}. \quad (2.1)$$

Here,  $x$  denotes the total synaptic current arriving at the soma, and  $y$  is the neuron’s firing rate. The parameters  $a$  and  $b$  change the slope and offset

of the sigmoid nonlinearity. In Triesch (2005b) we proposed an update rule for the parameters of the sigmoid based on the idea of moment matching. This rule took the form of two proportional control laws coupling the offset and slope of the sigmoid to estimates of the first and second moment of the neuron's firing-rate distribution.<sup>1</sup> More recently we derived a gradient rule that aims to directly minimize the Kullback-Leibler divergence between the neuron's current firing-rate distribution and the optimal exponential distribution (Triesch, 2005a). As derived in the appendix, this ansatz leads to the following discrete learning rule:  $a \leftarrow a + \Delta a$  and  $b \leftarrow b + \Delta b$ , with

$$\Delta a = \eta_{\text{IP}} \left( \frac{1}{a} + x - \left( 2 + \frac{1}{\mu} \right) xy + \frac{1}{\mu} xy^2 \right) \quad (2.2)$$

$$\Delta b = \eta_{\text{IP}} \left( 1 - \left( 2 + \frac{1}{\mu} \right) y + \frac{1}{\mu} y^2 \right), \quad (2.3)$$

where  $\mu$  is the neuron's desired mean activity and  $\eta_{\text{IP}}$  is a small learning rate. This rule is very similar to the one derived by Bell and Sejnowski (1995) for a single sigmoid neuron maximizing its entropy. However, our rule has additional terms stemming from the objective of keeping the mean firing rate low. Note that this rule is strictly local. The only quantities used to update the neuron's nonlinear transfer function are  $x$ , the total synaptic current arriving at the soma, and the firing rate  $y$ . Furthermore, it relies on only the neuron's current activity  $y$  and its squared activity  $y^2$ , which might be estimated from the cell's internal calcium concentration.

## 2.1 Experiments with the Intrinsic Plasticity model

*2.1.1 Behavior for Different Distributions of Total Synaptic Current.* To illustrate the behavior of the learning rule, we consider a single-model neuron with a fixed distribution of synaptic current  $f_x(x)$ . Figure 1 illustrates the result of adapting the neuron's nonlinear transfer function to three different input distributions: gaussian, uniform, and exponential. We set the unit's desired mean activity to  $\mu = 0.1$ , a tenth of its maximum activity, to reflect the low firing rates observed in cortex. The following results do not critically depend on the precise value of  $\mu$  as long as  $\mu$  is small ( $\mu \ll 1$ ). In each case, the sigmoid nonlinearity moves close to the optimal nonlinearity that would result in an exactly exponential distribution of the firing rate, resulting in a sparse distribution of firing rates. Note that since the sigmoid nonlinearity bounds activity to be less than one ( $y < 1$ ), the resulting distribution  $p_y(y)$  necessarily has a truncated tail. Also, since the sigmoid has

---

<sup>1</sup> In this work the sigmoid nonlinearity was parameterized slightly differently, with  $a$  representing the inverse slope of the sigmoid.

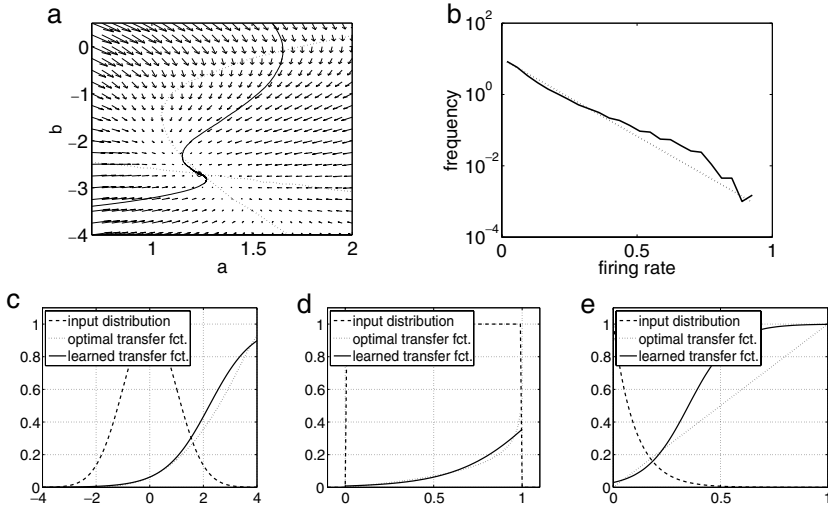


Figure 1: Dynamics of the IP rule for various input distributions. (a–c) Gaussian input distribution. (a) The phase plane diagram. Arrows indicate the flow field of the system. Dotted lines indicate approximate locations of the nullclines (found numerically). Two example trajectories are exhibited (solid lines), which converge to the stationary point (marked with a circle). (b) The resulting activity distribution (solid curve), which approximates that of an exponential distribution (dotted). (c) The theoretically optimal transfer function (dotted) and the learned sigmoidal transfer function (solid). The gaussian input distribution (dashed, not drawn to scale) is also shown. (d, e) Same as c but for uniform and exponential input distribution. Parameters were  $\mu = 0.1$ ,  $\eta_{IP} = 0.001$ .

only two degrees of freedom, the match is usually not perfect. As can be seen in the figure, large deviations from the optimal transfer function can sometimes be observed where the probability density of the input distribution is low, and this leads to deviations from the desired exponential activity distribution. A closer fit could be obtained with a different nonlinearity  $g$  with more free parameters. Since it is unclear, however, what additional degrees of freedom biological neurons have in this respect, we chose to stick with equation 2.1.

When a constant input  $x_0$  is presented to the unit— $f_x(x)$  is described by a  $\delta$ -function—then the slope parameter  $a$  will diverge as the unit attempts to adjust its nonlinearity to match its deterministic output  $g_{ab}(x_0)$  to the desired exponential distribution. Due to simultaneous adjustment of the threshold parameter  $b$ , however, the unit's activity will remain close to  $\mu$ .

The results obtained with equations 2.2 and 2.3 are very similar to the ones from our previous study, which used a set of two simple control laws

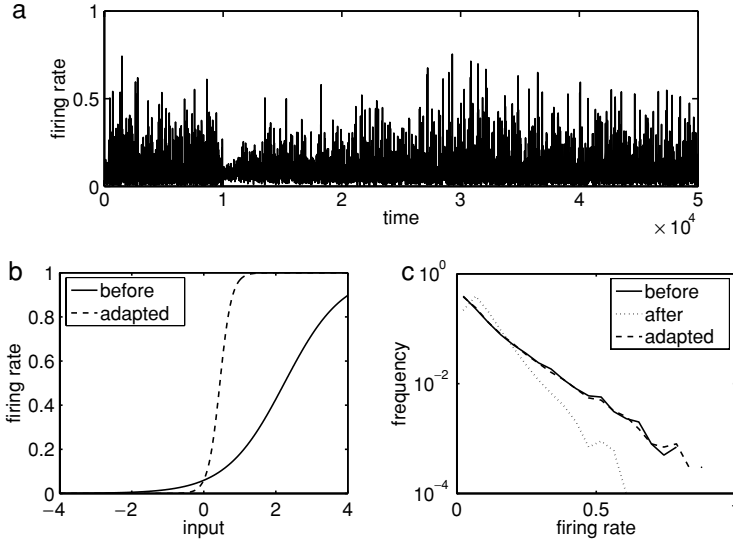


Figure 2: Response to sudden deprivation from afferent input. (a) The unit’s firing rate  $y$  as a function of the number of presented inputs (time  $t$ ). We plot the response only to every 10th input. After  $10^4$  time steps, the standard deviation of the input signal is reduced by a factor of five. In response, the neuron adjusts its intrinsic excitability to restore its desired exponential firing-rate distribution. (b) The nonlinear activation function before the deprivation and after adaptation to deprivation. (c) The distribution of firing rates before deprivation, immediately after deprivation, and after adaptation to deprivation.

that updated the sigmoid nonlinearity based on estimates of the first and second moment of the neuron’s firing-rate distribution (Triesch, 2005b). The current rule seems to lead to faster convergence, however.

**2.1.2 Response to Sudden Sensory Deprivation.** The proposed form of plasticity of a neuron’s intrinsic excitability may help neurons maintain stable firing-rate distributions in the presence of systematic changes to their afferent input. An interesting special case of such a change is that of sensory deprivation. Here, we model sensory deprivation as a fivefold reduction of the standard deviation  $\sigma$  of a gaussian-distributed total synaptic current  $x$  arriving at the soma. Figure 2 shows how the neuron learns to adjust to the sudden change of the distribution of total synaptic current occurring after  $10^4$  time steps. After a transient phase of low activity, the neuron gradually reestablishes its desired firing-rate level (see Figure 2a) by making its activation function steeper, that is, the neuron becomes more excitable. This is illustrated in Figure 2b, which shows the neuron’s activation function

before the reduction of afferent input and after adjusting to the new input distribution. Figure 2c shows how the distribution of firing rates is affected by the sudden reduction in afferent input and the subsequent adjustment of the neuron's excitability. After adjustment, the neuron regains its original, approximately exponential firing-rate distribution.

**2.2 Discussion of the IP model.** It is interesting to note that equation 2.1 can be viewed as linearly transforming input  $x$  before passing it through a standard sigmoid function. Such a linear rescaling could also be obtained by modifying all synaptic weights (including a "bias" weight) entering the neuron. Thus, there is a redundancy between the neuron's intrinsic and synaptic degrees of freedom. This "redundancy" of neural degrees of freedom in our model is specific to our choice of the nonlinearity and its parameterization. In fact, the derivation of the learning rule can be generalized to other nonlinearities with different adjustable parameters. The only requirements are that the nonlinearity  $g$  should be strictly monotonically increasing and differentiable with respect to  $x$ , and the partial derivatives of  $\log \partial y / \partial x$  with respect to the parameters of  $g$  (in the above case,  $a$  and  $b$ ) must exist.

Nevertheless, there is ample evidence that homeostatic synaptic mechanisms also contribute to keeping a neuron's activity in a desirable regime (Turrigiano & Nelson, 2000, 2004), especially if the afferent input to the neuron has been reduced (Maffei, Nelson, & Turrigiano, 2004). From a self-organization viewpoint, however, it may be easiest and most effective to implement mechanisms that change the spiking statistics of the neuron directly at the soma—as locally as possible. But since such mechanisms may operate over only a limited range, they may rely on other processes such as synaptic scaling to keep the distribution of currents arriving at the soma in a desired range. In fact, most recent evidence suggests that such synaptic scaling is sensitive to glutamate levels in the extracellular medium (Stellwagen & Malenka, 2006), which is more directly related to the synaptic input  $x$  to the neuron rather than its firing activity. Thus, the brain may use synaptic scaling to keep the level of input to a neuron in a desired regime while using intrinsic mechanisms to map this input to its own firing activity in a nearly optimal way.

Changes in the slope and threshold of a neuron's  $f - I$  curve as well as combinations of both have been observed experimentally. Such changes occur through the modification of voltage-gated channels including, for example,  $\text{Na}^+$  channels (Aizenman, Akerman, Jensen, & Hollis, 2003),  $\text{K}^+$  channels (van Welie et al., 2006), and  $I_h$  conductances (van Welie et al., 2004). Many of such modifications seem to depend on the cell's internal  $\text{Ca}^{2+}$  concentration, which is often regarded as an indicator of the neuron's activity level. Modeling studies have also revealed that concerted changes of  $I_A$  and  $I_h$  could implement the kind of gain changes associated with modification of the  $a$ -parameter in our parametric sigmoid (Burdakov, 2005). In sum,

it appears plausible that changes to the  $f - I$  function as predicted by our model could be implemented through the modification of voltage-gated channels.

The most important prediction of our IP model(s) for experiments is that changes to the intrinsic excitability of a neuron should not be driven only by the average spiking activity of the neuron—the first moment of the neuron’s activity—but also depend on higher moments. In particular, our gradient rule and our earlier rule based on moment matching (Triesch, 2005b) rely on the second moment of the neuron’s activity. This implies that experimental manipulations that leave a neuron’s average activity constant but change the second moment of its activity will lead to systematic changes in the neuron’s  $f - I$  curve. The required estimate of the second moment of the neuron’s firing rate may be implemented by an agent  $A$  that binds two  $\text{Ca}^{2+}$  ions:  $A + 2\text{Ca}^{2+} \rightarrow A'$ . The concentration of  $A'$  could then approximate the square of the current firing rate.

Furthermore, we hypothesize that the functional role of homeostatic synaptic plasticity is to maintain the synaptic current arriving at the soma in a desired regime, while the role of IP is to map this distribution of synaptic current onto the desired distribution of spiking activity. In principle, this hypothesis can be tested by any experimental technique that allows independently controlling a neuron’s synaptic input and its firing activity.

### 3 Interaction of IP with Hebbian Learning

---

Since IP appears to be a ubiquitous phenomenon in the brain, it is important to study how it interacts with other forms of plasticity, in particular with associative synaptic learning mechanisms. We begin by considering a standard Hebbian learning rule,

$$\Delta \mathbf{w} = \eta_{\text{Hebb}} \mathbf{u} y(\mathbf{u}) = \eta_{\text{Hebb}} \mathbf{u} g_{ab}(\mathbf{w}^T \mathbf{u}), \quad (3.1)$$

where  $\mathbf{u}$  is the vector of synaptic inputs to the neuron and  $\eta_{\text{Hebb}}$  is a small learning rate. Note that since  $y$  is positive,  $\Delta w_i$  will always be positive if  $u_i$  is positive. Thus, for positive inputs  $u_i$ , the strength of weight  $w_i$  can only increase. Such uncontrolled weight growth is undesirable, even if the IP counteracts this effect and stabilizes the neuron’s activity. In order to avoid unlimited weight growth, we keep the weight vector fixed at unit length through the multiplicative weight normalization  $\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\|$ . Future work could also incorporate a model of synaptic scaling to keep the input to the neuron in a desired regime (see the discussion in section 2). Since synaptic scaling also seems to function in a multiplicative fashion, (e.g., Turrigiano, Leslie, Desai, Rutherford, & Nelson, 1998), we expect that this would lead to similar results.

**3.1 Behavior in the Limit of Fast IP.** It is instructive to consider the limiting case of fast IP:  $\eta_{IP} \gg \eta_{Hebb}$ . In this case, the IP will have generated an approximately exponential distribution of  $y(\mathbf{u})$ , before the weight vector  $\mathbf{w}$  can change much. Note that our discussion is independent of the exact IP mechanism (e.g., the specific learning rule introduced above) as long as the IP mechanism is effective in producing approximately exponential firing-rate distributions. In the case of fast IP, it can be seen from equation 3.1 that the expected value of the weight update given by

$$E[\Delta \mathbf{w}] = \eta_{Hebb} E[\mathbf{u}y(\mathbf{u})]$$

is an exponentially weighted sum of all the inputs  $\mathbf{u}$ , with the firing rates  $y(\mathbf{u})$  functioning as the exponentially distributed weighting factors. A small number of inputs that produce the highest firing rates will exert a strong pull on the weight vector, while the majority of inputs lead to a small firing rate and change the weight vector only marginally. This mechanism can lead to the discovery of heavy-tailed directions in the input, as illustrated by the following experiment. A careful analysis for clustered inputs is given below.

We consider a model neuron with just two inputs. The input distribution is such that there is one heavy-tailed direction and an orthogonal direction without heavy tails. At the same time, the distribution is white—that is, the covariance matrix of the inputs is the  $2 \times 2$  identity matrix. For concreteness we consider the Laplace band whose pdf is given by

$$p(u_1, u_2) = p(u_1)p(u_2) = \begin{cases} \frac{1}{2\sqrt{6}} \exp(-\sqrt{2}|u_1|) & : |u_2| \leq \sqrt{3} \\ 0 & : |u_2| > \sqrt{3} \end{cases} \quad (3.2)$$

Along the  $u_1$ -direction, this distribution has the shape of a Laplacian with exponential tails, while along the  $u_2$ -direction, this distribution is uniform with no tails at all (see Figure 3a). Note that since this distribution has the identity matrix as its covariance matrix, standard Hebbian learning in a linear unit would perform a random walk in the weight space; it cannot discover the heavy-tailed direction. In combination with IP, however, the weight vector is drawn toward the heavy-tailed direction (see Figures 3b and 3c; parameters were  $\mu = 0.1$ ,  $\eta_{IP} = 0.01$ ,  $\eta_{Hebb} = 0.001$ ). This process can be understood as follows. Consider the weight vector in an oblique orientation (the thin arrow in Figure 3b). For this weight vector, we plot a sample of inputs (circles) with their sizes representing how strongly they will activate the unit. The inputs leading to the highest firing rates (biggest circles) are more likely to appear to the right of the weight vector, at the far right of the input distribution. This imbalance will cause the weight vector to turn clockwise in the direction of the heavy tail. This process ends when

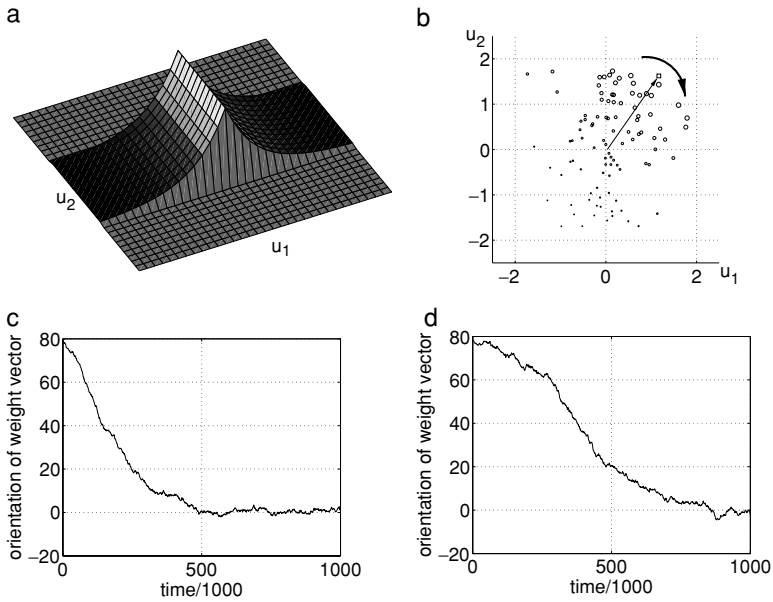


Figure 3: Discovery of heavy-tailed direction in a model neuron with just two inputs. See the text for details.

the weight vector has aligned with the heavy-tailed  $u_1$ -direction, which corresponds to an orientation of zero degrees in Figure 3c.

We have also experimented with a number of other input distributions and found that the property of seeking out the more heavy-tailed directions is a robust effect. Figure 3d demonstrates this for the case of a white input distribution that is the product of a gaussian along the  $u_2$  direction and a Laplacian along the  $u_1$  direction. The weight vector aligns with the more heavy-tailed  $u_1$  direction (parameters as above). Note that while we have assumed IP being much faster than Hebbian learning, the same behavior is also exhibited if IP is slower than Hebbian learning. This is a robust effect that holds true in this example and for all experiments discussed in this article (an example is given in Figure 8).

This mechanism for finding heavy-tailed directions is closely related to a number of approaches for independent component analysis (ICA) that use nonlinear Hebbian terms in their learning rules. A prominent example is the FastICA algorithm (Hyvärinen & Oja, 1997; Hyvärinen, Karhunen, & Oja, 2001). Since FastICA works with a fixed nonlinearity, it is not surprising that our method finds heavy-tailed direction even when IP is slow. However, one of the benefits of IP is that it drives the neuron toward using an energy-efficient code.

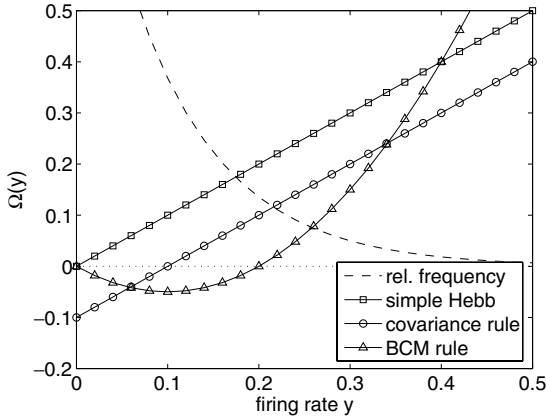


Figure 4: Illustration of alternative Hebbian learning rules, plotting the functions  $\Omega(y)$  for three different learning rules. For the covariance and BCM rules, we plot  $\Omega$  for the thresholds that balance LTP and LTD. For the BCM rule,  $\Omega$  has been scaled by a factor five for better visibility. The dotted horizontal line marks the boundary between LTP ( $\Omega(y) > 0$ ) and LTD ( $\Omega(y) < 0$ ). The dashed line shows the relative probability of firing rate  $y$  (not drawn to scale).

**3.2 Alternative Hebbian Learning Rules.** We now discuss the interaction of IP with alternative Hebbian learning rules. We focus on generalizations of the Hebbian learning rule, equation 3.1, of the kind

$$\Delta \mathbf{w} = \eta_{\text{Hebb}} \mathbf{u} \Omega(y(\mathbf{u})), \quad (3.3)$$

where  $\Omega$  is a generally nonlinear function. The choice  $\Omega_{\text{id}}(y) = y$  corresponds to the standard Hebbian learning rule from above. Other choices lead to different learning rules, as illustrated in Figure 4.

The choice  $\Omega_{\text{cov}}(y) = y - \theta_{\text{cov}}$  is a so-called covariance rule, since it can lead to the discovery of the principal eigenvector of the input covariance matrix in a linear unit (e.g., Dayan & Abbott, 2001). To this end,  $\theta_{\text{cov}}$  is set to be an online estimate of the unit's mean activity. The choice  $\Omega_{\text{BCM}}(y) = (y - \theta_{\text{BCM}})y$  is a so-called quadratic BCM rule (Cooper, Intrator, Blais, & Shouval, 2004), which automatically stabilizes weight growth. Here,  $\theta_{\text{BCM}}$  is usually an online estimate of the unit's mean squared firing rate. In our analysis, we will keep the thresholds  $\theta_{\text{cov}}$  and  $\theta_{\text{BCM}}$  fixed, since the IP drives the unit to a fixed firing-rate distribution automatically.

Both the covariance and the BCM rule differ from the standard rule by introducing long-term depression (LTD) in addition to long-term potentiation (LTP). In connection with IP as introduced above, it is easy to relate the amount of LTP versus LTD to the thresholds used in the learning rules.

A natural choice for the thresholds is one that balances LTP and LTD. The specific condition for this balance is that  $\Omega(y)$  is on average zero:<sup>2</sup>

$$\int_0^\infty \frac{1}{\mu} \exp\left(-\frac{y}{\mu}\right) \Omega(y) dy = 0. \quad (3.4)$$

In the case of the covariance rule, this condition leads to  $\theta_{\text{cov}} = \mu$ , that is, LTD is exhibited for inputs that lead to below-average firing rates, and LTP is exhibited for above-average firing rates (see Figure 4). For the BCM rule, equation 3.4 leads to  $\theta_{\text{BCM}} = 2\mu$ . Thus, in this balanced LTP-LTD regime, the BCM rule will exhibit maximal LTD for inputs that correspond to the mean firing rate and LTP occurs only for inputs that lead to firing rates greater than twice the average firing rate (compare Figure 4). Firing rates in the tail of the exponential distribution have a very strong influence on the weight vector due to the quadratic increase of  $\Omega(y)$ .

Empirically, we found that such alternative Hebbian learning rules are also effective in finding heavy-tailed directions when combined with our IP mechanism. Based on the arguments on the relation of nonlinear Hebbian learning rules to ICA algorithms discussed above, this is not surprising. In particular, it has been argued that BCM-type learning rules can discover sparse directions in the absence of any IP (Cooper et al., 2004).

**3.3 IP as an Alternative to the Sliding Threshold in BCM Theory.** A hallmark of the BCM theory of synaptic plasticity is the idea of a sliding threshold that separates postsynaptic firing rates producing LTP versus LTD (Cooper et al., 2004). LTP occurs only when the postsynaptic activity is above the threshold  $\theta_{\text{BCM}}$ , but  $\theta_{\text{BCM}}$  is assumed to be changing as a function of the cell's activity history. The evidence for such a sliding LTP-LTD threshold seems to be only indirect, however, and IP may present a plausible alternative to the sliding threshold proposed in BCM learning.

One of the pieces of evidence put forward as support for the idea of a sliding threshold is data from Kirkwood and colleagues demonstrating that for dark-reared cats, weaker inputs to a neuron are sufficient to produce LTP when compared to normally reared cats (Kirkwood, Rioult, & Bear, 1996). The threshold stimulation frequency that separates LTD and LTP is smaller for the dark-reared cats, which is consistent with the idea of a sliding threshold on the postsynaptic activity level. But an alternative explanation of these data is that neurons in the dark-reared cats are more excitable such that the same stimulation frequency leads to higher firing levels, and this is

---

<sup>2</sup> This may not be the only sensible choice, however. Another form of balance would result from requiring that half of the inputs induce LTP, while the other half induce LTD; that is, we require  $\Omega(\mu \ln(2)) = 0$ , where  $\mu \ln(2)$  is the median firing activity of the unit. This choice leads to  $\theta_{\text{cov}} = \theta_{\text{BCM}} = \mu \ln(2)$ .

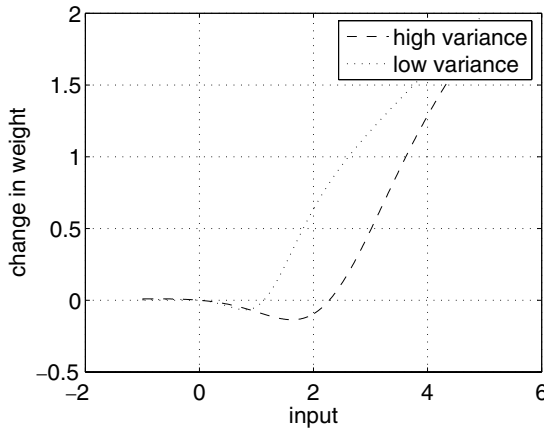


Figure 5: Intrinsic plasticity may provide an alternative explanation to the idea of a sliding threshold as introduced in the BCM theory of Hebbian plasticity. The change of the weight  $\Delta w$  is plotted as a function of the strength of the input driving the neuron. For a neuron that experiences four times lower variance inputs, the plasticity curve shifts to the left because IP makes the neuron more excitable. Due to this, the unit in the low-variance input environment will exhibit a stronger response and thus a stronger change in weight for the same input.

why weaker inputs can already induce LTP, even for a fixed modification threshold. Figure 5 illustrates this effect. It shows a plasticity curve, where the change of the weight  $\Delta w$  corresponding to the amount of LTD or LTP is plotted as a function of the strength of the input to the neuron. We plot this for two neurons that have received different levels of stimulation in the past and have adjusted their activation function accordingly. Both units receive white zero mean gaussian input of different variance. The unit receiving input with four times lower variance corresponds to the situation of the dark-reared cats. Due to the low variance input, it has assumed a steeper activation function that produces higher outputs for the same input, as shown in Figure 2. As a consequence, its plasticity function is shifted to the left. Thus, shifts in the plasticity function can in principle be explained by IP and do not require the assumption of a sliding threshold. This mechanism may work in conjunction with homeostatic synaptic scaling mechanisms, as mentioned earlier.

**3.4 Analysis for Clustered Inputs.** A more formal analysis of the interaction between IP and Hebbian learning rules can be performed when we assume that the input distribution is a mixture of  $N$  well-separated clusters in a high-dimensional input space. Let the cluster centers be at locations  $\mathbf{c}_i, i = 1, \dots, N$ . We treat only the case where the clusters have equal

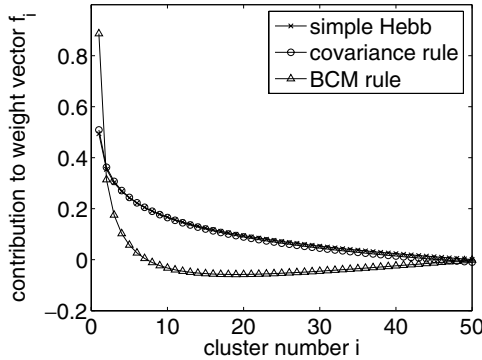


Figure 6: Behavior of three Hebbian learning rules for clustered input. The contribution that each input cluster makes to the weight vector is plotted as a function of the cluster number.

prior probabilities. The extension to the general case is straightforward. As shown in the appendix, for the case of the standard Hebbian learning rule, the stationary solutions for the weight vector  $\mathbf{w}$  will be a weighted sum of the cluster centers  $\mathbf{c}_i$ ,

$$\mathbf{w} = \sum_{i=1}^N f_i^{\text{Hebb}} \mathbf{c}_i, \tag{3.5}$$

where the weighting factors  $f_i^{\text{Hebb}}$  are given by

$$f_i^{\text{Hebb}} \propto 1 + \log(N) - i \log(i) + (i - 1) \log(i - 1), \tag{3.6}$$

and we define  $0 \log(0) \equiv 0$ . This relation is plotted in Figure 6. Here, the clusters have been ordered such that cluster 1 is the one closest to the final weight vector and cluster  $N$  is the one most distant from the final weight vector. There can be at most  $N!$  resulting weight vectors owing to the number of possible assignments of the  $f_i^{\text{Hebb}}$  to the clusters, but in general not all of them are stationary points of the learning dynamics. An interesting result of this analysis is that the final weight vector does not depend on the desired mean activity level  $\mu$ . In the special case of the cluster centers lying at random locations in a very high-dimensional space, all of the possible stationary weight vectors correspond to heavy-tailed directions. More precisely, the linear projection of the inputs onto the weight vector will have an approximately exponential distribution, that is, its distribution will have a heavier tail than a gaussian (Triesch, 2005b). This formal result complements the informal analysis from above, which

suggested that the combination of IP with Hebbian learning leads to the discovery of heavy-tailed directions in the input.

We have performed the same analysis for the case of the alternative Hebb rules introduced above (see the appendix). For the covariance rule, the contributions of the individual clusters to the weight vector are very similar to the contributions in the simple Hebbian case, as shown in Figure 6, although a few of the weights will actually become slightly negative. For the BCM rule, however, most of the input clusters will contribute to the weight vector with a small negative weight and only a small fraction of the clusters is contributing to the final weight vector with a strong positive weight. The exact size of this fraction depends on the choice of the threshold  $\theta_{\text{BCM}}$ . Figure 6 shows the result for the “balanced” choice  $\theta_{\text{BCM}} = 2\mu$ .

#### 4 The Bars Problem

---

The bars problem is a standard problem for unsupervised learning architectures (Földiák, 1990). Formally, it is a nonlinear independent component analysis problem. The input domain consists of an  $N$ -by- $N$  retina. On this retina, all horizontal and vertical bars ( $2N$  in total) can be displayed. The presence or absence of each bar is determined independently, with every bar occurring with the same probability  $p$  (in our case,  $p = 1/N$ ). If a horizontal and a vertical bar overlap, the pixel at the intersection point will be just as bright as any other pixel on the bars rather than twice as bright. This makes the problem a nonlinear ICA problem. Example stimuli from the bars data set are shown in Figure 7a. Note that we normalize input vectors to unit length. The goal of learning in the bars problem is to find the independent sources of the images: the individual bars. Thus, the neural learning system should develop filters that represent the individual bars.

We have trained an individual sigmoidal model neuron on the bars input domain. We used the simple Hebbian rule, equation 3.1, together with the multiplicative weight normalization that keeps the weight vector at unit length. The theoretical analysis above assumed that intrinsic plasticity is much faster than synaptic plasticity. Here, we set the timescale of IP to be identical to the one for synaptic plasticity ( $\eta_{\text{IP}} = 0.01$ ,  $\eta_{\text{Hebb}} = 0.01$ ). As illustrated in Figure 7b, the unit’s weight vector aligns with one of the individual bars as soon as the intrinsic plasticity has pushed the model neuron into a regime where its responses are sparse: the unit has discovered one of the independent sources of the input domain. This result is robust if the desired mean activity  $\mu$  of the unit is changed over a wide range. If  $\mu$  is reduced from its default value ( $1/2N = 0.05$ ) over several orders of magnitude (we tried down to  $10^{-5}$ ), the result remains unchanged. However, if  $\mu$  is increased above about 0.15, the unit will fail to represent an individual bar but will learn a mixture of two or more bars, with different bars being represented with different strengths. Thus, in this example—in contrast to the theoretical result above—the desired mean activity  $\mu$  does influence the

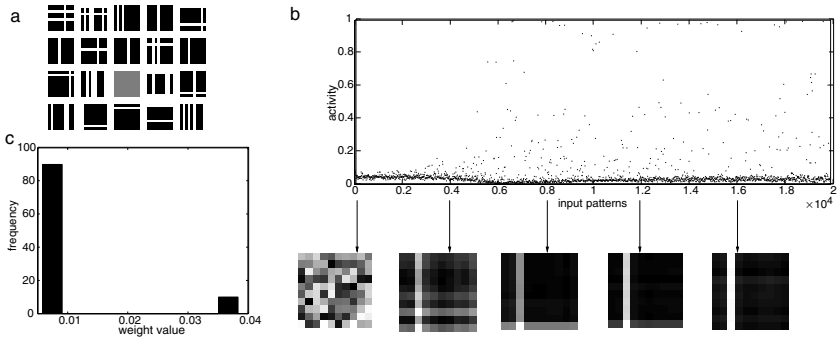


Figure 7: Single-model neuron discovering a “bar” in the bars problem. (a) Examples of bars stimuli. (b) Activity of the neuron during the learning process and weight vector at five times during learning. (c) Histogram of weight strength of the final weight vector. Parameters were  $\eta_{IP} = 0.01$ ,  $\eta_{Hebb} = 0.01$ .

weight vector that is being learned. The reason is that the IP only imperfectly adjusts the output distribution to the desired exponential shape. As long as  $\mu$  is small, however, discovery of a single bar is a very robust effect. For example, it still occurs if all the input patterns presented to the system have at least, say, four bars in them (i.e., the unit never sees a single bar in isolation).

The discovery of a single bar also does not critically depend on the relative time scales of IP and Hebbian learning. Figure 8 shows results for different relative speeds of IP and Hebbian learning. When IP is faster than Hebbian learning (left graph,  $\eta_{IP} = 0.01$ ,  $\eta_{Hebb} = 0.001$ ), a single bar is discovered as soon as the unit’s activity is driven into a heavy-tailed regime. When IP is slower than Hebbian learning (center graph,  $\eta_{IP} = 0.001$ ,  $\eta_{Hebb} = 0.01$ ), the same behavior is observed, although convergence is slower.

Importantly, this behavior is not observed when IP is absent, as shown in Figure 8 (right). In this experiment, we used a fixed sigmoid nonlinearity with  $a = 5.0$ ,  $b = -1.15$ . These values correspond to the final sigmoid after convergence in Figure 7.<sup>3</sup> With this fixed nonlinearity, the unit is unable to discover a bar. We varied the learning rate  $\eta_{Hebb}$  over two orders of magnitude and also ran longer simulations, but the unit never discovered a single bar. This demonstrates that IP contributes more to the learning process than just setting the sigmoid nonlinearity to its final shape. Instead, it shows that the interaction of IP with Hebbian learning over an extended time contributes to the final result. This extended interaction leads to large transient

<sup>3</sup>More precisely, in the “converged state,” these parameters still fluctuate.  $a$  stays mostly in the interval [4.5, 5.5].  $b$  fluctuates in the interval [−1.0, −1.3].

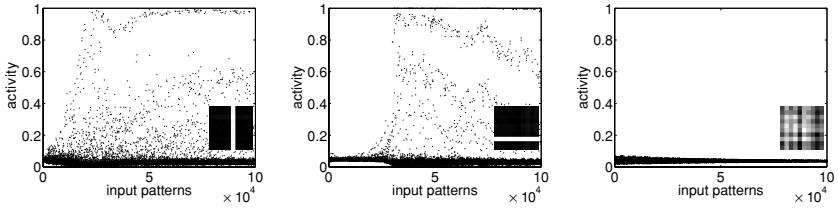


Figure 8: Learning is robust to changes in the relative timescale of IP versus Hebbian learning. Faster IP (left plot,  $\eta_{IP} = 0.01$ ,  $\eta_{Hebb} = 0.001$ ) and slower IP (center plot,  $\eta_{IP} = 0.001$ ,  $\eta_{Hebb} = 0.01$ ) both lead to the discovery of a single bar (see final weight vectors shown as small insets). Without any IP (right plot,  $\eta_{Hebb} = 0.001$ ), but with a constant nonlinearity that corresponds to the final nonlinearity in Figure 7, no bar is discovered.

deviations of the sigmoid from its final shape. During this transient deformation, the unit's activity first enters into a sparse regime and discovers a bar. This weight pattern is maintained as the unit keeps adjusting its nonlinearity to the final shape. It should be noted, however, that setting  $a$  and  $b$  to constant values that are assumed during the transient phase can also lead to successful discovery of a bar. Thus, IP is not strictly necessary for successful learning, but it automatically adjusts the nonlinearity in the desired way and ensures an energy-efficient, approximately exponential activity distribution throughout the learning process.

## 5 Discussion

We have presented a model of IP rooted in information theory and have shown that the model is effective in maintaining homeostasis of the neuron's firing-rate distribution by making it approximately exponential. The model predicts that changes of the neuron's intrinsic excitability should depend not only on the neuron's average firing rate (first moment or mean) but also on its average squared firing rate (second moment). This mechanism may complement other homeostasis mechanisms such as synaptic scaling (Turrigiano & Nelson, 2000, 2004; Maffei et al., 2004). Recent evidence suggests that homeostatic synaptic scaling may occur in response to changing levels of afferent input to the cell rather than to changes of the cell's firing activity (Stellwagen & Malenka, 2006). We propose that synaptic scaling tries to keep the synaptic current arriving at the soma in a certain regime, while intrinsic plasticity adjusts the neuron's nonlinear properties to map this current onto an energy-efficient, sparse firing pattern.

**5.1 Connection to Posttraumatic Epileptogenesis.** We demonstrated that the proposed IP mechanism contributes to homeostasis by increasing

a neuron's excitability after reduction of its afferent input. This effect may play an important role in the genesis of posttraumatic epilepsy. Physiological studies have found increased excitability of pyramidal neurons after trauma (Prince & Tseng, 1993), and modeling studies have shown that such increased excitability together with increased NMDA conductance may explain posttraumatic epileptogenesis (Bush, Prince, & Miller, 1999). Our IP model suggests a causal connection between the trauma and the increased excitability. Due to the trauma, neurons in the neighboring tissue lose a fraction of their afferent input, which triggers an increased excitability due to IP. This in turn makes the local circuit more prone to generating epileptiform events. This mechanism may be complementary to the putative role of homeostatic synaptic plasticity in posttraumatic epileptogenesis (Houweling, Bazhenov, Timofeev, Steriade, & Sejnowski, 2005). More work is needed to understand the interaction of these mechanisms in detail, with potential implications for clinical applications.

**5.2 Role for Learning Efficient Sensory Representations.** We have demonstrated that there may be a synergistic relationship between Hebbian synaptic learning and IP mechanisms that changes the intrinsic excitability of neurons. The two processes may work together to find heavy-tailed directions in the input space. While our theoretical arguments explained this behavior only in the limiting case of IP being much faster than synaptic plasticity, we found empirically that the effect persists if intrinsic plasticity is as slow as or even much slower than synaptic plasticity. In addition, we have demonstrated how IP may mimic the effect of a sliding threshold for synaptic plasticity as postulated by the BCM theory. New experiments are needed to distinguish between the alternative explanations of a sliding plasticity threshold or a change in excitability with a fixed threshold.

Our work is closely related to other work on finding efficient representations of sensory stimuli in general (Olshausen & Field, 1996; Bell & Sejnowski, 1997) and on doing so with biologically plausible plasticity mechanisms in particular. For example, some authors have used modified versions of standard Hebbian learning rules in order to discover independent components in the input (Blais, Intrator, Shouval, & Cooper, 1998; Hyvärinen & Oja, 1998), and certain independent component analysis techniques also rely on terms that multiply the input signal with a nonlinearly transformed output signal (Hyvärinen & Oja, 1997; Hyvärinen et al., 2001). Thus, our work is closely related to these approaches, but it introduces IP as a mechanism to adapt the nonlinearity online in a biologically viable fashion. To our knowledge, this was ignored in previous work on the topic. Our result from Figure 8 (right), suggests that at least in certain instances, this online adaptation aids learning. In addition, the IP mechanism ensures an energy-efficient encoding throughout the learning process.

**5.3 Coupling with Reward Mechanisms.** Good sensory representations must efficiently encode the aspects of the input domain that are relevant for the organism's behavior, which implies not wasting neural resources on representing irrelevant information. While in the experiments on the bars problem, a model neuron learned to represent a single bar at random, it is easy to envision how a simple gating mechanism can lead to learning of relevant heavy-tailed directions in a biologically plausible fashion (e.g., Montague, Dayan, Person, & Sejnowski, 1995; Roelfsema & van Ooyen, 2005). For example, the following modification of the synaptic learning rule allows the discovery of a specific bar that is associated with rewards or punishments:

$$\Delta w = \eta y x R. \quad (5.1)$$

In this equation,  $R$  represents a scalar relevance signal identifying situations that may be worth learning about. For example, we may define  $R \equiv |\delta|$ , where  $\delta$  represents a temporal difference error (Sutton & Barto, 1998) which has been associated with the activity of dopaminergic neurons. The use of  $|\delta|$  reflects that both positive and negative events may be worth learning about. If, for example, an unexpected positive or negative reward is given ( $\delta = \pm 1$ ), whenever a specific bar appears (irrespective of the presence or absence of any other bars), the modified learning rule will discover this specific bar, because the gating signal effectively restricts learning to those stimuli that contain this bar. Such a relevance feedback could be the effect of dopamine, as suggested above, but it could also be associated with the action of acetylcholine, which is known to globally influence synaptic plasticity (Pennartz, 1995). It would be interesting to incorporate the ideas presented in this article into an actor-critic framework to study how the emergence of efficient sensory representations may be influenced by task demands.

**5.4 Future Work.** As a next step, it is important to study the effect of combined intrinsic and Hebbian synaptic plasticity at the network level. While we have shown that intrinsic plasticity can effectively ensure the lifetime sparseness of individual neurons (Willmore & Tolhurst, 2001) in the sense of exponential firing-rate distributions, additional lateral inhibitory interactions in a network are likely necessary to decorrelate units in a population in order to also ensure population sparseness, as is observed in the cortex (Vinje & Gallant, 2000). We speculate that the cerebral cortex utilizes a synergistic interaction of intrinsic and synaptic plasticity mechanisms within a sophisticated system of local excitation and inhibition to find sensory codes that exhibit both lifetime and population sparseness (Lehky, Sejnowski, & Desimone, 2005). Preliminary work has already established that topographic arrangements of simple cell-like receptive fields emerge in a two-dimensional network of interconnected units with simultaneous intrinsic and synaptic learning (Butko & Triesch, 2006).

It will also be interesting to study the effects of combining intrinsic and synaptic learning mechanisms in recurrent network models. Recently we have also started to investigate the interaction of IP models for spiking neurons with spike-timing-dependent plasticity (STDP) in recurrent networks and found that the combination of STDP and IP can produce very different network dynamics from those resulting from STDP without IP. We speculate that similar combinations of plasticity mechanisms may be effective in driving recurrent networks into a regime of great computational power (Bertschinger & Natschläger, 2004) for a fixed energy expenditure.

In conclusion, since IP appears to be a ubiquitous phenomenon in the brain, many aspects of neural functioning and plasticity could be profoundly influenced by it. A careful analysis of the interaction of various forms of plasticity in the nervous system is arguably vital for our understanding of the brain's ability to adapt to ever changing demands during normal development and learning and in response to various disturbances such as trauma or infection. Our analysis of the interaction of Hebbian learning with the proposed IP model is a small step in this direction. In a similar vein, Janowitz and van Rossum (2006) have explored the role of a fast nonhomeostatic IP mechanism for trace conditioning. So far, however, we have seen only a small glimpse of this terra incognita.

## Appendix A: Gradient Rule for IP

---

We describe the distribution of the total synaptic current arriving at the soma  $x$  by the probability distribution  $f_x(x)$ . Assuming the neuron's nonnegative firing rate is given by  $y = g(x)$ , where  $g$  is strictly monotonically increasing, the distribution of  $y$  is given by

$$f_y(y) = \frac{f_x(x)}{\frac{\partial y}{\partial x}}. \quad (\text{A.1})$$

We are looking for ways to adjust  $g$  so that  $f_y(y)$  is approximately exponential, that is, we would like  $f_y(y)$  to be "close" to  $f_{\text{exp}}(y) = \frac{1}{\mu} \exp\left(\frac{-y}{\mu}\right)$ . A natural measure is to consider the Kullback-Leibler divergence (KL divergence) between  $f_y$  and  $f_{\text{exp}}$ :

$$D \equiv d(f_y \| f_{\text{exp}}) = \int f_y(y) \log \left( \frac{f_y(y)}{\frac{1}{\mu} \exp\left(\frac{-y}{\mu}\right)} \right) dy \quad (\text{A.2})$$

$$= \int f_y(y) \log(f_y(y)) dy - \int f_y(y) \left( -\frac{y}{\mu} - \log \mu \right) dy \quad (\text{A.3})$$

$$= -H(y) + \frac{1}{\mu} E(y) + \log \mu, \quad (\text{A.4})$$

where  $H(y)$  is the entropy of  $y$  and  $E(y)$  is its expected value. Thus, in order to minimize  $d(f_y \| f_{\text{exp}})$ , we need to maximize the entropy  $H(y)$  while minimizing the expected value  $E(y)$ . The term  $\log \mu$  does not depend on  $g$  and is irrelevant for this minimization.

To derive a stochastic gradient-descent rule for IP that strives to minimize equation A.4, we need to take into account the specific form of the parameterized nonlinearity, equation 2.1. To construct the gradient, we calculate the partial derivatives of  $D$  with respect to  $a$  and  $b$ :

$$\frac{\partial D}{\partial a} = \frac{\partial d(f_y \| f_{\text{exp}})}{\partial a} = -\frac{\partial H}{\partial a} + \frac{1}{\mu} \frac{\partial E(y)}{\partial a} \quad (\text{A.5})$$

$$= E \left( -\frac{\partial}{\partial a} \log \left( \frac{\partial y}{\partial x} \right) + \frac{1}{\mu} \frac{\partial y}{\partial a} \right) \quad (\text{A.6})$$

$$= -\frac{1}{a} + E \left( -x + \left( 2 + \frac{1}{\mu} \right) xy - \frac{1}{\mu} xy^2 \right), \quad (\text{A.7})$$

where we have used equation A.1 and moved the differentiation inside the expected value operation to obtain equation A.6 and exploited

$$\log \left( \frac{\partial y}{\partial x} \right) = \log a + \log y + \log(1 - y) \quad (\text{A.8})$$

in the last step. Similarly, for the partial derivative with respect to  $b$ , we find:

$$\frac{\partial D}{\partial b} = \frac{\partial d(f_y \| f_{\text{exp}})}{\partial b} = -\frac{\partial H}{\partial b} + \frac{1}{\mu} \frac{\partial E(y)}{\partial b} \quad (\text{A.9})$$

$$= -1 + E \left( \left( 2 + \frac{1}{\mu} \right) y - \frac{1}{\mu} y^2 \right). \quad (\text{A.10})$$

The resulting stochastic gradient-descent rule is given by  $a \leftarrow a + \Delta a$  and  $b \leftarrow b + \Delta b$ , with:

$$\Delta a = \eta_{\text{IP}} \left( \frac{1}{a} + x - \left( 2 + \frac{1}{\mu} \right) xy + \frac{1}{\mu} xy^2 \right) \quad (\text{A.11})$$

$$\Delta b = \eta_{\text{IP}} \left( 1 - \left( 2 + \frac{1}{\mu} \right) y + \frac{1}{\mu} y^2 \right). \quad (\text{A.12})$$

**Appendix B: Analysis of Learning with Clustered Inputs**

We consider the limiting case of fast IP and we assume that IP is perfect—it leads to an exactly exponential distribution of activities (Triesch, 2005b). We will first treat the case of only two input clusters and then generalize to  $N$  clusters.

Let us assume two clusters of inputs at locations  $\mathbf{c}_1$  and  $\mathbf{c}_2$ . Let us also assume that both clusters account for exactly half of the inputs. If the weight vector is slightly closer to one of the two clusters, inputs from this cluster will activate the unit more strongly. Let  $m = \mu \ln(2)$  denote the median of the exponential firing-rate distribution with mean  $\mu$ . Then inputs from the closer cluster, say  $\mathbf{c}_1$ , will be responsible for all activities above  $m$ , while the inputs from the other cluster will be responsible for all activities below  $m$ . Hence, for the standard Hebb rule, the expected value of the weight update  $\langle \Delta \mathbf{w} \rangle$  will be given by:

$$\langle \Delta \mathbf{w} \rangle \approx \alpha \mathbf{c}_1 \int_m^\infty \frac{y}{\mu} \exp(-y/\mu) dy + \alpha \mathbf{c}_2 \int_0^m \frac{y}{\mu} \exp(-y/\mu) dy \quad (\text{B.1})$$

$$= \frac{\alpha \mu}{2} ((1 + \ln 2) \mathbf{c}_1 + (1 - \ln 2) \mathbf{c}_2). \quad (\text{B.2})$$

Taking the multiplicative weight normalization into account, we see that stationary points of the weight vector require that  $\Delta \mathbf{w}$  is parallel to  $\mathbf{w}$ . From this, it follows that the weight vector must converge to either of the following two stationary states:

$$\mathbf{w} = \frac{(1 \pm \ln 2) \mathbf{c}_1 + (1 \mp \ln 2) \mathbf{c}_2}{\|(1 \pm \ln 2) \mathbf{c}_1 + (1 \mp \ln 2) \mathbf{c}_2\|}. \quad (\text{B.3})$$

The weight vector moves close to one of the two clusters but does not fully commit to it.

In the case of  $N$  clusters, the situation is similar. In order to calculate the contribution of the  $i$ th closest cluster to the weight vector, we have to consider the inverse of the cumulative distribution function (percent point function) of the exponential firing-rate distribution, which is given by

$$F_y^{-1}(p) = -\mu \log(1 - p). \quad (\text{B.4})$$

(Recall that the median of a random variable  $y$  is given by  $F_y^{-1}(1/2)$ .) Assuming that all clusters contribute the same number of inputs, the cluster associated with the  $i$ th highest interval of firing rates is “responsible” for the interval  $[1 - \frac{i}{N}, 1 - \frac{i-1}{N}]$  of the cumulative distribution function. It

follows that the average amount of weight update associated with inputs from cluster  $i$  under the standard Hebb rule is given by

$$\int_{F_y^{-1}(1-\frac{i}{N})}^{F_y^{-1}(1-\frac{i-1}{N})} y \frac{1}{\mu} \exp\left(-\frac{y}{\mu}\right) dy, \quad (\text{B.5})$$

which reduces to

$$\mu \left[ \frac{i}{N} \left( 1 - \log\left(\frac{i}{N}\right) \right) - \frac{i-1}{N} \left( 1 - \log\left(\frac{i-1}{N}\right) \right) \right]. \quad (\text{B.6})$$

This expression determines the relative contribution of cluster  $i$  to the weight vector in the steady state. After simplification, we can see that possible stable weight vectors are of the form

$$\mathbf{w} = \sum_{i=1}^N f_i^{\text{Hebb}} \mathbf{c}_i, \quad (\text{B.7})$$

where

$$f_i^{\text{Hebb}} \propto 1 + \log(N) - i \log(i) + (i-1) \log(i-1) \quad (\text{B.8})$$

is the relative contribution of the  $i$ th cluster to the weight vector. Note that for the case of  $N = 2$ , this matches the result above (see equation B.3).

It is straightforward to generalize this analysis to the alternative Hebbian learning rules introduced in the main text. In order to calculate the expected weight update due to cluster  $i$ , we generalize equation B.5 to

$$\int_{F_y^{-1}(1-\frac{i}{N})}^{F_y^{-1}(1-\frac{i-1}{N})} \Omega(y) \frac{1}{\mu} \exp\left(-\frac{y}{\mu}\right) dy. \quad (\text{B.9})$$

For the balanced covariance rule  $\Omega(y) = y - \mu$ , we obtain

$$f_i^{\text{cov}} \propto f_i^{\text{Hebb}} - \frac{\mu}{N}. \quad (\text{B.10})$$

For the balanced BCM rule  $\Omega(y) = y(y - 2\mu)$ , we obtain:

$$f_i^{\text{BCM}} \propto i \left( \log \frac{i}{N} \right)^2 - (i-1) \left( \log \frac{i-1}{N} \right)^2. \quad (\text{B.11})$$

In Figure 6 we plot normalized  $f_i$  such that  $\sum_i f_i^2 = 1$ .

## Acknowledgments

---

This work was supported in part by the National Science Foundation under grant NSF IIS-0208451 and the Hertie foundation. I thank Nicholas Butko, Emanuel Todorov, and Erik Murphy-Chutorian for discussions and Cornelius Weber and two anonymous reviewers for helpful comments on earlier drafts.

## References

---

- Aizenman, C., Akerman, C., Jensen, K., & Hollis, T. (2003). Visually driven regulation of intrinsic neuronal excitability improves stimulus detection in vivo. *Neuron*, *39*, 831–842.
- Baddeley, R., Abbott, L. F., Booth, M., Sengpiel, F., & Freeman, T. (1998). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. London, Ser. B*, *264*, 1775–1783.
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*, 1129–1159.
- Bell, A. J., & Sejnowski, T. J. (1997). The “independent components” of natural scenes are edge filters. *Vision Research*, *37*, 3327–3338.
- Bertschinger, N., & Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, *16*, 1413–1436.
- Bienenstock, E., Cooper, L., & Munro, P. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, *2*, 32–48.
- Blais, B. S., Intrator, N., Shouval, H., & Cooper, L. N. (1998). Receptive field formation in natural scene environments. *Neural Computation*, *10*, 1797–1813.
- Burdakov, D. (2005). Gain control by concerted changes in  $i_a$  and  $i_h$  conductances. *Neural Computation*, *17*, 991–995.
- Bush, P., Prince, D., & Miller, K. (1999). Increased pyramidal excitability and NMDA conductance can explain posttraumatic epileptogenesis without disinhibition: A model. *Journal of Neurophysiology*, *82*(4), 1748–1758.
- Butko, N. J., & Triesch, J. (2006). Exploring the role of intrinsic plasticity for the learning of sensory representations. In M. Verleysen (Ed.), *European Symposium on Artificial Neural Networks (ESANN)* (pp. 467–472). Bruges, Belgium: d-side.
- Cooper, L., Intrator, N., Blais, B., & Shouval, H. (2004). *Theory of cortical plasticity*. River Edge, NJ: World Scientific Publishing.
- Cudmore, R., & Turrigiano, G. (2004). Long-term potentiation of intrinsic excitability in LV visual cortical neurons. *J. Neurophysiol.*, *92*, 341–348.
- Daoudal, G., & Debanne, D. (2003). Long-term plasticity of intrinsic excitability: Learning rules and mechanisms. *Learning and Memory*, *10*, 456–465.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT Press.
- Desai, N. S., Rutherford, L. C., & Turrigiano, G. G. (1999). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature Neuroscience*, *2*(6), 515–520.
- Földiák, P. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, *64*, 165–170.

- Houweling, A., Bazhenov, M., Timofeev, I., Steriade, M., & Sejnowski, T. (2005). Homeostatic synaptic plasticity can explain post-traumatic epileptogenesis in chronically isolated neocortex. *Cerebral Cortex*, *15*, 834–845.
- Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: Wiley.
- Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, *9*, 1483–1492.
- Hyvärinen, A., & Oja, E. (1998). Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, *64*(3), 301–313.
- Janowitz, M., & van Rossum, M. (2006). Excitability changes that complement Hebbian learning. *Network: Computation in Neural Systems*, *17*(1), 31–41.
- Kirkwood, A., Rioult, M., & Bear, M. (1996). Experience-dependent modification of synaptic plasticity in visual cortex. *Nature*, *381*, 526–528.
- Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.*, *36c*, 910–912.
- Lehky, S., Sejnowski, T., & Desimone, R. (2005). Selectivity and sparseness in the responses of striate complex cells. *Vision Research*, *45*, 57–73.
- Maffei, A., Nelson, S., & Turrigiano, G. (2004). Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation. *Nature Neuroscience*, *7*(12), 1353–1359.
- Marder, E., Abbott, L. F., Turrigiano, G. G., Liu, Z., & Golowasch, J. (1996). Memory from the dynamics of intrinsic membrane currents. *Proc. Natl. Acad. Sci. USA*, *93*, 13481–13486.
- Montague, P., Dayan, P., Person, C., & Sejnowski, T. (1995). Bee foraging in uncertain environments using predictive Hebbian learning. *Nature*, *377*, 725–728.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.
- Pennartz, C. (1995). The ascending neuromodulatory systems in learning by reinforcement: Comparing computational conjectures with experimental findings. *Brain Res. Rev.*, *21*, 219–245.
- Prince, D., & Tseng, G.-F. (1993). Epileptogenesis in chronically injured cortex: In vitro studies. *Journal of Neurophysiology*, *69*, 1276–1291.
- Roelfsema, P. R., & van Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural Computation*, *17*, 2176–2214.
- Shin, J., Koch, C., & Douglas, R. (1999). Adaptive neural coding dependent on the time-varying statistics of the somatic input current. *Neural Computation*, *11*, 1893–1913.
- Stellwagen, D., & Malenka, R. C. (2006). Synaptic scaling mediated by glial TNF- $\alpha$ . *Nature*, *440*(7087), 1054–1059.
- Stemmler, M., & Koch, C. (1999). How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nature Neuroscience*, *2*(6), 521–527.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
- Triesch, J. (2005a). A gradient rule for the plasticity of a neuron's intrinsic excitability. In W. Duch, J. Kacprzyk, E. Oja, & S. Zadrozny (Eds.), *Proc. Int. Conf. on Artificial Neural Networks (ICANN)* (pp. 65–70). Berlin: Springer.

- Triesch, J. (2005b). Synergies between intrinsic and synaptic plasticity in individual model neurons. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems*, 17. Cambridge, MA: MIT Press.
- Turrigiano, G., Leslie, K., Desai, N., Rutherford, L., & Nelson, S. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391, 892–896.
- Turrigiano, G., & Nelson, S. (2000). Hebb and homeostasis in neuronal plasticity. *Curr. Opin. Neurobiol.*, 10(3), 358–364.
- Turrigiano, G., & Nelson, S. (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 5, 97–107.
- van Welie, I., van Hooft, J. A., & Wadman, W. J. (2004). Homeostatic scaling of neuronal excitability by synaptic modulation of somatic hyperpolarization-activated  $I_h$  channels. *Proc. Nat. Acad. Sci.*, 101(14), 5123–5128.
- van Welie, I., van Hooft, J. A., & Wadman, W. J. (2006). Background activity regulates excitability of rat hippocampal CA1 pyramidal neurons by adaptation of a  $K^+$  conductance. *J. Neurophysiology*, 95, 2007–2012.
- Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287, 1273–1276.
- Willmore, B., & Tolhurst, D. (2001). Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems*, 12, 255–270.
- Zhang, M., Hung, F., Zhu, Y., Xie, Z., & Wang, J.-H. (2004). Calcium signal-dependent plasticity of neuronal excitability developed postnatally. *J. Neurobiol.*, 61(2), 277–287.
- Zhang, W., & Linden, D. J. (2003). The other side of the engram: Experience-driven changes in neuronal intrinsic excitability. *Nature Reviews Neuroscience*, 4, 885–900.