Generating functionals for computational intelligence: The Fisher information as an objective function for self-limiting Hebbian learning rules

Rodrigo Echeveste [1]**, Claudius Gros** [1]

[1] *Institute for Theoretical Physics, Goethe University Frankfurt, Germany.*

Correspondence*:
Claudius Gros
Goethe University Frankfurt, Germany, gros07@itp.uni-frankfurt.de

ABSTRACT

Generating functionals may guide the evolution of a dynamical system and constitute a possible route for handling the complexity of neural networks as relevant for computational intelligence. We propose and explore a new objective function which allows to obtain plasticity rules for the afferent synaptic weights. The adaption rules are Hebbian, self-limiting, and result from the minimization of the the Fisher information with respect to the synaptic flux.

We perform a series of simulations examining the behavior of the new learning rules in various circumstances. The vector of synaptic weights aligns with the principal direction of input activities, whenever one is present. A linear discrimination is performed when there are two or more principal directions; directions having bimodal firing-rate distributions, being characterized by a negative excess kurtosis, are preferred.

We find robust performance and full homeostatic adaption of the synaptic weights results as a by-product of the synaptic flux minimization. This self-limiting behavior allows for stable online learning for arbitrary durations. The neuron acquires new information when the statistics of input activities is changed at a certain point of the simulation, showing however a distinct resilience to unlearn previously acquired knowledge. Learning is fast when starting with randomly drawn synaptic weights and substantially slower when the synaptic weights are already fully adapted.

Keywords: Hebbian learning   generating functionals   synaptic plasticity   objective functions   Fisher information   homeostatic adaption

1 INTRODUCTION

Synaptic plasticity involves the modification of the strength of individual synapses as a function of pre- and postsynaptic neural activity. Hebbian plasticity (**Hebb**, 2002) tends to reinforce already strong synapses and may hence lead, on a single neuron level, to runaway synaptic growth, which needs to be contained through homeostatic regulative processes (**Turrigiano and Nelson**, 2000), such as synaptic scaling (**Abbott and Nelson**, 2000). Modeling of these dual effects has been typically a two-step approach, carried out by extending Hebbian-type learning rules by regulative scaling principles (**Bienenstock et al.**, 1982; **Oja**, 1992; **Goodhill and Barrow**, 1994; **Elliott**, 2003).

An interesting question regards the fundamental computational task a single neuron should be able to perform. There is a general understanding that synaptic scaling induces synaptic competition and that this synaptic competition generically results in a generalized principal component analysis (PCA) (**Oja**, 1992; **Miller and MacKay**, 1994), in the sense that a neuron will tend to align its vector of synaptic weights, within the space of input activities, with the direction having the highest variance. A meaningful behavior, since information possibly transmitted by input directions with low variances is more susceptible to be obfuscated by internal or environmental noise.

A single neuron may however have additional computational capabilities, in addition to its basic job as a principal component analyzer. The neuron may try to discover 'interesting directions', in the spirit of projection pursuit (**Huber**, 1985), whenever the covariance matrix of the afferent inputs is close to unity. Deviations from Gaussian statistics may encode in this case vitally important information, a well known feature of natural image statistics (**Simoncelli and Olshausen**, 2001; **Sinz and Bethge**, 2013). One measure for non-Gaussianess is given by the kurtosis (**DeCarlo**, 1997) and a single neuron may possibly tend to align its synaptic weight vector with directions in the space of input activities characterized by heavy tails (**Triesch**, 2007), viz having a large positive excess kurtosis. Here we study self-limiting Hebbian plasticity rules which allow the neuron to discover maximally bimodal directions in the space of input activities, viz directions having a large negative excess kurtosis.

Binary classification in terms of a linear discrimination of objects in the input data stream is a basic task for neural circuits and has been postulated to be a central component of unsupervised object recognition within the framework of slow feature analysis (**Wiskott and Sejnowski**, 2002; **DiCarlo et al.**, 2012). It is of course straightforward to train, using supervised learning rules, a neuron to linearly separate the data received into two categories. Here we propose that a single neuron may perform this task unsupervised, whenever it has a preference for directions in the space of input activities characterized by negative excess kurtosis. Neural signals in the brain containing high frequency bursts have been linked to precise information transmission (**Lisman**, 1997). Neurons switching between relatively quiet and bursting states tend to have bimodal firing rate distributions and negative excess kurtosis. The autonomous tendency to perform a binary classification, on a single neuron level, may hence be of importance for higher cortical areas, as neurons would tend to focus their intra-cortical receptive fields towards intermittent bursting neural populations. A subclasss of bursting pyramidal neurons have been found in layer 5 of somatosensory and visual cortical areas (**Chagnac-Amitai et al.**, 1990). Neurons receiving input from these bursting cortical neurons would therefore be natural candidates to test this hypothesis, for which there is, to date, no direct experimental evidence.

In order to develop synaptic plasticity rules, one may pursue one of two routes: either to reproduce certain aspects of experimental observations by directly formulating suitable plasticity rules, or to formulate, alternatively, an objective function from which adaption rules are then deduced (**Intrator and Cooper**, 1992; **Bell and Sejnowski**, 1995). Objective functions, also denoted generating functionals in the context of dynamical system theory (**Linkerhand and Gros**, 2013a; **Gros**, 2014), generically facilitate higher-level investigations, and have been used, e.g., for such as an overall stability analysis of Hebbian-type learning in autonomously active neural networks (**Dong and Hopfield**, 1992).

The Fisher information measures the sensitivity of a system with respect to a given parameter. It can be related, in the context of population coding (**Brunel and Nadal**, 1998), to the transfer information between stimulus and neural activity, and to order-parameter changes within the theory of thermodynamic phase transitions (**Prokopenko et al.**, 2011). Minimization of the Fisher information can be used as a generative principle for quantum mechanics in general (**Reginatto**, 1998) and for the Euler equation in density functional theory (**Nagy**, 2003). Here we propose an objective function for synaptic learning rules based on the Fisher information with respect to a differential operator we denote the synaptic flux.

The aim of adapting synaptic weights is to encode a maximal amount of information present in the statistics of the afferent inputs. The statistics of the output neural activity becomes stationary when this task is completed and the sensitivity of the activity of the post-synaptic neuron with regard to changes in the synaptic weights is then minimal. Minimizing the Fisher information with respect to the synaptic flux

is hence a natural way to generate synaptic plasticity rules. Morover, as we show in section 3, the synaptic plasticity rules obtained by minimizing the Fisher information for the synaptic flux have a set of attractive features; incorporating standard Hebbian updating and being, at the same time, self-limiting.

Minimizing an information theoretical objective function, like the Fisher information, is an instance of polyhomeostatic optimization (**Marković and Gros**, 2010), namely the optimization of an entire function. Other examples of widely used information theoretical measures are the transfer entropy (**Vicente et al.**, 2011) and the Kullback-Leibler divergence, which one may use for adapting, on a slow time scale, intrinsic neural parameters like the bias, also called offset (**Triesch**, 2007; **Marković and Gros**, 2012). Minimizing the Kullback-Leibler divergence then corresponds to maximizing the information content, in terms of Shannon's information entropy, of the neural firing rate statistics. We use intrinsic adaption for self-regulating the bias, obtaining, as a side effect, an effective sliding threshold for the synaptic learning rule, in spirit of the BCM rule (**Bienenstock et al.**, 1982).

2   MATERIALS AND METHODS

In the present work we consider rate encoding neurons for which the output firing rate y is obtained as a sigmoidal function of the membrane potential x via:

$$y = \sigma(x - b), \qquad \sigma(z) = \frac{1}{1 + e^{-z}}, \qquad x = \sum_{j=1}^{N_w} w_j (y_j - \bar{y}_j) , \tag{1}$$

where N_w is the number of input synapses, and w_j and y_j represent the synaptic weights and firing rates of the afferent neurons respectively. The sigmoidal $\sigma(z)$ has a fixed gain (slope) and the neuron has a single intrinsic parameter, the bias b. The $\bar{y}_j$ represent the trailing averages of y_j,

$$\frac{d}{dt}\bar{y}_j = \frac{y_j - \bar{y}_j}{T_y} , \tag{2}$$

with T_y setting the time scale for the averaging. Synaptic weights may take, for rate encoding neurons, both positive and negative values and we assume here that afferent neurons firing at the mean firing rate $y_j \simeq \bar{y}_j$ do not influence the activity of the postsynaptic neuron. This is a standard assumption for synaptic plasticity which is incorporated in most studies by appropriately shifting the mean of the input distribution.

In what follows we will derive synaptic plasticity rules for the w_j and intrinsic plasticity rules that will optimize the average magnitude of x and set in this way, implicitly, the gain of the transfer function. We have not included an explicit gain acting on x since any multiplicative constant can be absorbed into the w_j and, conversely, the average value of the w_j can be thought of as the gain of the transfer function with rescaled w_j.

The firing rate y of neurons has an upper limit, an experimental observation which is captured by restricting the neural output of rate encoding neurons to the range $y \in [0, 1]$. Here we consider with

$$F_{ob}(x, y) \;=\; E\left[\left(2 + x\,(1 - 2y)\,\right)^2\right] \tag{3}$$

an objective function for synaptic plasticity rules which treats the upper and the lower activity bounds on an equal footing. $E[\cdot]$ denotes the expectation value.

The functional F_{ob} is positive definite and can be expressed as in (3), or purely as a function of either x or $y = \sigma(x - b)$. In Fig. 1, F_{ob} is plotted as a function of y for different values of the bias b. The functional always presents two minima and diverges for extremal firing rates $0/1$. In particular, for firing rates $y \to 0/1$, F_{ob} is minimized by membrane potentials $x \to (-2)/2$, respectively.
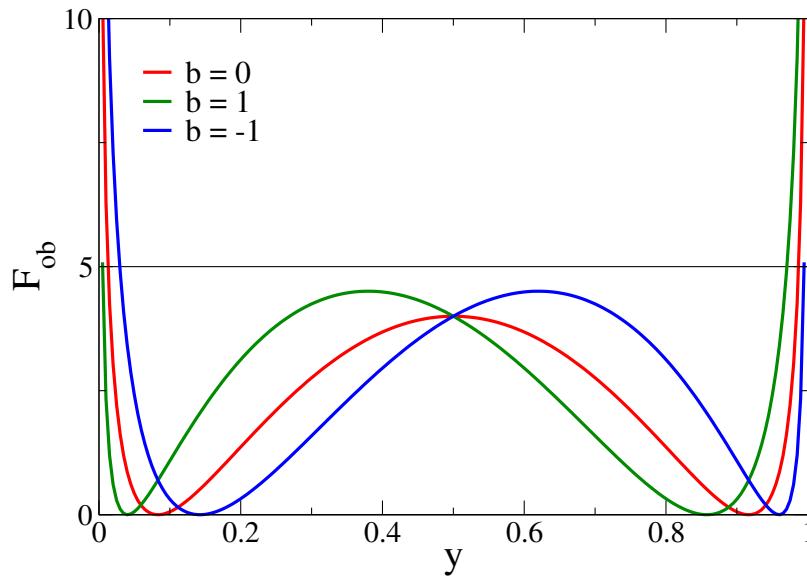
Figure 1. The objective function F_{ob}, expression (3), as a function of the output firing rate y for different values of the bias b. F_{ob} always has two minima and diverges for extremal firing rates $y \to 0/1$, a feature responsible for inducing output firing rates.

Minimizing (3) as an objective function for deriving synaptic plasticity rules will therefore lead to bounded membrane potentials and hence necessarily to bounded learning rules, devoid of runaway synaptic growth. The cost function (3) generically has two distinct minima, a feature setting it apart from other objective functions for synaptic plasticity rules (**Intrator and Cooper**, 1992). Moreover, the objective function (3) can also be motivated by considering the Fisher information of the postsynaptic firing rate with respect to the synaptic flux, as shown in section 2.1.

In section 2.2, via stochastic gradient descent, the following plasticity rules for the afferent synaptic weights w_j will be derived:

$$\dot{w}_j \;=\; \epsilon_w G(x) H(x)(y_j - \bar{y}_j), \tag{4}$$

with

$$G(x) = 2 + x(1 - 2y), \qquad H(x) = (2y - 1) + 2x(1 - y)y. \tag{5}$$

Here ϵ_w controls the rate of the synaptic plasticity. The bias b entering the sigmoidal may be either taken to be constant or adapted via

$$\dot{b} \;=\; (-\epsilon_b)\big(1 - 2y + y(1 - y)\lambda\big), \tag{6}$$

in order to obtain a certain average postsynaptic firing rate, where λ is a control parameter, as detailed out in section 2.2. Eq. (6) leads to the optimization of the statistical information content of the neural activity, in terms of Shannon's information entropy; a process also denoted intrinsic adaption (**Triesch**, 2007) or polyhomeostatic optimization (**Marković and Gros**, 2010).

Both adaption rules, (4) for the synaptic plasticity and (6) for regulating the average postsynaptic firing rate, interfere only weakly. For instance one could take the bias b as a control parameter, by setting $\epsilon_b \to 0$, and measure the resulting mean firing rate a posteriori. The features of the synaptic adaption process remain unaffected and therefore alternative formulations for the intrinsic adaption of the bias could also be considered.

The synaptic plasticity rule (4) involves the Hebbian factor $H(x)$, and a multiplicative synaptic weight rescaling factor $G(x)$. Although here G and H are presented as a function of x and y, these can also be expressed entirely in terms of y, consistently with the Hebbian interpretation. It is illustrative to consider

the cases of small/large postsynaptic neural activity. In the limit $y \to 0/1$, which is never reached, the updating rules (4) would read

$$
\dot{w}_j \;\propto\; \left\{ \begin{array}{ll} (2+x)\,(-1)\,(y_j - \bar{y}_j) & (y \to 0) \\ (2-x)\,(+1)\,(y_j - \bar{y}_j) & (y \to 1) \end{array} \right. . \tag{7}
$$

For the case that $|x| < 2$ we hence have that the synaptic strength decreases/increases for an active presynaptic neuron with $y_j > \bar{y}_j$, whenever the postsynaptic neuron is inactive/active, an instance of Hebbian learning. The multiplicative constraint $(2 \pm x)$ in (7) results in a self-limitation of synaptic growth. Synaptic potentiation is turned into synaptic depression whenever the drive x becomes too large in magnitude. Runaway synaptic growth is hence not possible and the firing rate will settle close to the minima of F_{ob}, compare Fig. 1.

2.1 MOTIVATION IN TERMS OF FISHER INFORMATION

The synaptic plasticity rules (4) can be derived either directly from the objective function (3), as explained in section 2.2 or motivated from an higher-order principle, the optimization of the synaptic flux, as we will show in the following. Synaptic weight competition could be formulated, as a matter of principles, through an ad-hoc constraint like

$$
\sum_j (w_j)^2 \to \text{const.}, \qquad \mathbf{w} = (w_1, w_2, \dots) , \tag{8}
$$

which defines a hypersphere in the phase of afferent synaptic weights $\{w_j\}$, together with some appropriate Hebbian-type adaption rules. We will not make use of (8) explicitly, but our adaption rules implicitly lead to finite length for the synaptic weight vector $\mathbf{w}$.

Synaptic plasticity will modify, quite generically, the statistical properties of the distribution $p(y)$ of the firing rate y of the postsynaptic neuron. It is hence appropriate to consider the sensitivity of the firing-rate distribution $p(y)$ with respect to changes in the w_j. For this purpose one may make use of the Fisher information

$$
F_\theta \;=\; \int p(y) \left(\frac{\partial}{\partial \theta} \ln \big(p(y) \big) \right)^2 dy , \tag{9}
$$

which encodes the sensitivity of a given probability distribution function $p(y)$ with respect to a certain parameter θ. Here we are interested in the sensitivity with respect to changes in the synaptic weights $\{w_j\}$ and define with

$$
F_w \;=\; \int p(y) \left(\sum_j w_j \frac{\partial}{\partial w_j} \ln \big(p(y) \big) \right)^2 dy , \tag{10}
$$

the Fisher information with respect to the synaptic flux. Expression (10) corresponds to the Fisher information (9) when considering

$$
\frac{\partial}{\partial \theta} \;\to\; \sum_j w_j \frac{\partial}{\partial w_j} = \mathbf{w} \cdot \nabla_w, \tag{11}
$$

as differential operator. The factors w_j in front of the $\partial/\partial w_j$ result in a dimensionless expression, the generating functional (10) is then invariant with respect to an overall rescaling of the synaptic weights and the operator (11) a scalar. Alternatively we observe that $w_j \partial/\partial w_j = \partial/\partial \log(w_j/w_0)$, where w_0 is an arbitrary reference synaptic weight, corresponding to the gradient in the space of logarithmically discounted synaptic weights.
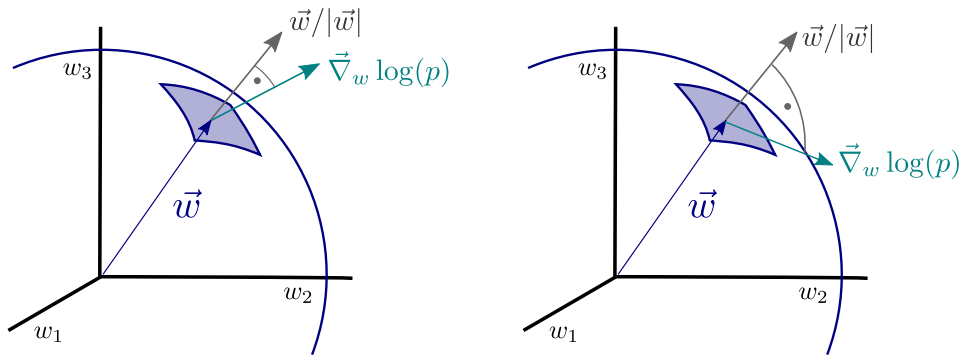
Figure 2. Illustration of the principle of minimal synaptic flux. The synaptic flux, compare expression (11), is the scalar product between the gradient $\nabla_{\mathbf{w}} \log(p)$ and the normal vector of the synaptic sphere, $\vec{w}/|\vec{w}|$ (left). Here we disregard the normalization. The sensitivity $\nabla_{\mathbf{w}} \log(p)$ of the neural firing-rate distribution $p = p(y)$, with respect to the synaptic weights $\mathbf{w} = (w_1, w_2, w_3, \dots)$, vanishes when the local synaptic flux is minimal (right), viz when $\mathbf{w} \cdot \nabla_{\mathbf{w}} \log(p) \to 0$. At this point the magnitude of the synaptic weight vector $\vec{w}$ will not grow anymore.

The operator (11), which we denote *synaptic flux operator*, is, in addition, invariant under rotations within the space of synaptic weights and the performance of the resulting synaptic plasticity rules will hence be also invariant with respect to the orientation of the distributions $p(y_j)$ of the input activities $\{y_j\}$. Physically, the operator (11) corresponds, apart from a normalization factor, to the local flux through the synaptic hypersphere, as defined by Eq. (8), since the synaptic vector $\mathbf{w}$ is parallel to the normal vector through the synaptic hypersphere, as illustrated in Fig. 2.

The Fisher information (10) can be considered as a generating functional to be minimized with respect to the synaptic weights $\{w_j\}$. The time-averaged properties of the neural activities, as measured by $p(y)$, will then not change any more at optimality; the sensitivity of the neural firing-rate distribution, with respect to the synaptic weights, vanishing for small F_w. At this point the neuron has finished encoding the information present in the input data stream through appropriate changes of the individual synaptic weights.

It is interesting to consider what would happen if one would maximize the Fisher information instead of minimizing it. Then the neural firing activity would become very sensitive to small changes in the synaptic weights $\{w_j\}$ and information processing unstable, being highly susceptible to noise, viz to small statistical fluctuation of the synaptic weights. On a related note, the inverse Fisher information constitutes, via the Cramer-Rao theory (**Paradiso**, 1988; **Seung and Sompolinsky**, 1993; **Gutnisky and Dragoi**, 2008), a lower bound for the variance when estimating an external parameter θ. In this context, the external parameter θ can be estimated more reliably when the Fisher information is larger, viz when the distribution considered is highly sensible to the parameter of interest. This is a different setup. Here we are not interested in estimating the value of the synaptic weights, but in deducing adaption rules for the $\{w_j\}$.

2.2 SYNAPTIC FLUX MINIMIZATION

We are interested in synaptic plasticity rules which are instantaneous in time, depending only on the actual pre- and postsynaptic firing rates y_j and y. Hence, the actual minimization of the synaptic flux functional (10) needs to be valid for arbitrary distributions $p(y_j)$ of the presynaptic firing activities $\{y_j\}$. The synaptic flux F_w, which is in the first place a functional of the postsynaptic activity $p(y)$, needs therefore to be reformulated in terms of the distributions $p(y_j)$. A faithful representation of the postsynaptic firing-rate distribution entering F_w would involve a convolution over all presynaptic $p(y_j)$ and would hence lead to intricate cross-synaptic correlations (**Bell and Sejnowski**, 1995). Our aim here, however, is to develop synaptic plasticity rules for individual synapses, functionally dependent only on the local presynaptic

activity and on the overall postsynaptic firing level. We hence consider for the minimization of the synaptic flux all $j \in \{1, \ldots, N_w\}$ synapses separately, viz we replace (10) by

$$
\begin{aligned}
F_w \quad &\rightarrow \quad \int \left(\sum_j w_j \frac{\partial}{\partial w_j} \ln \left(\frac{p(y_j)}{\partial y / \partial y_j} \right) \right)^2 \prod_l p(y_l) dy_l \\
&\equiv \quad \int f_w(y) \prod_l p(y_l) dy_l \, ,
\end{aligned} \tag{12}
$$

where we have defined the kernel $f_w(y)$. We denote the approximation (12) the *local synapse approximation*, since it involves the substitution of $p(y)dy$ by $\prod_l p(y_l)dy_l$. Expression (12) becomes exact for the case $N_w = 1$. It corresponds to the case in which the distinct afferent synapses interact only via the overall value of the membrane potential x, as typical for a mean-field approximation. We then find, using the neural model (1),

$$
\frac{\partial y}{\partial y_j} \; = \; y(1-y)w_j \tag{13}
$$

and hence

$$
f_w(y) \; = \; \left(\sum_j w_j \left(\frac{1}{w_j} + (y_j - \bar{y}_j)(1 - 2y) \right) \right)^2 \; = \; \Big(N_w + x(1 - 2y) \Big)^2 \tag{14}
$$

where N_w is the number of afferent synapses. The kernel f_w is a function of y only, and not of the individual y_j, since $x = \sum_j w_j(y_j - \bar{y}_j)$. More fundamentally, this dependency is a consequence of choosing the flux operator (11) to be a dimensionless scalar.

Taking $N_w \rightarrow 2$ in (14) leads to the objective function (3) and results in $G(x)$ and $H(x)$ being proportional to each other's derivatives, with the roots and maxima respectively aligned. We however also performed simulations using the generic expression (14), with the results changing only weakly and quantitatively.

The synaptic weights are updated so that $f_w(y)$ becomes minimal, $\dot{w}_j \propto -\partial f_w(y)/\partial w_j$, obtaining the plasticity rule (4). This procedure corresponds to a stochastic steepest descent of the objective function (3), a procedure employed when one is interested in obtaining update rules independent of the actual distributions $p(y_j)$ of the afferent neural activities.

For the derivation one makes use of $\partial y/\partial w_j = (y_j - \bar{y}_j)(1 - y)y$. The synaptic plasticity rule (4) depends via $y_j - \bar{y}_j$ on the activity y_j of the presynaptic neuron relatively to its mean firing rate $\bar{y}_j$. This dependence models experimental findings indicating that, in the context of spike timing dependent plasticity, low-frequency stimulation generically induces causal depression (**Shouval et al.**, 2010; **Lisman and Spruston**, 2010; **Feldman**, 2012); one needs above-average firing rates for causal potentiation.

Note that synaptic competition is present implicitly in the updating rule through the membrane potential x, entering both $G(x)$ and $H(x)$, which integrates all individual contributions; the local synapse approximation (12) only avoids explicit cross synaptic learning.

We denote the two factors on the right-hand side of (4), $G(x)$ and $H(x)$, as self-limiting and Hebbian respectively; with $H(x)$ being, by construction, the derivative of $G(x)$. The derivative of $H(x)$ is also proportional to $G(x)$ since we substituted $N_w \rightarrow 2$ in the objective function on the right-hand-side of Eq. (14). With this choice, the two factors $G(x)$ and $H(x)$ are hence conjugate to each other.

The synaptic plasticity rule (4) works robustly for a wide range of adaption rates ϵ_w, including the case of online learning with constant updating rates. For all simulations presented here we have used
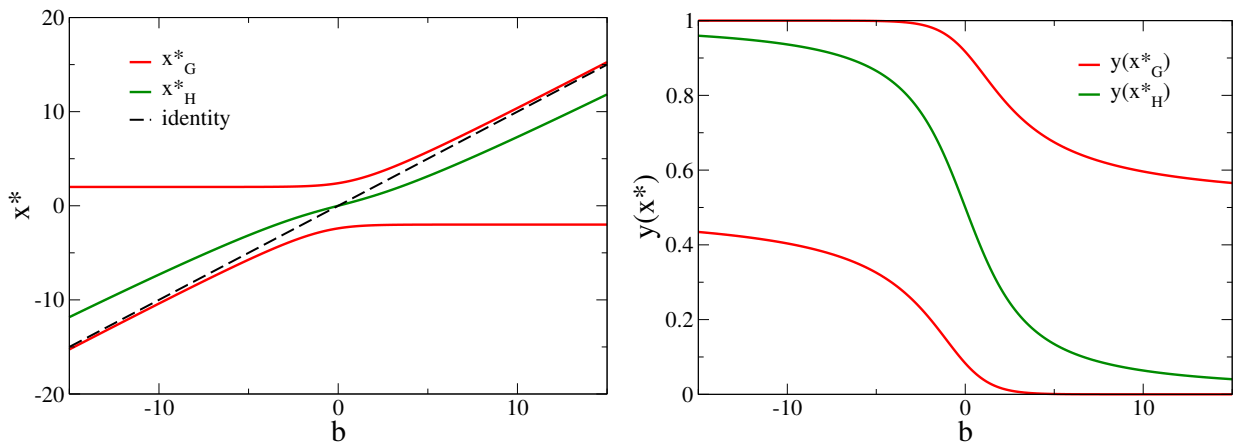
Figure 3. The roots of the adaption factors. Left: The roots $G(x_{0,1}^*) = 0$ and $H(x^*) = 0$ respectively, compare Eqs. (5) and (5), as a function of the bias b. Note that the roots do not cross, as the factors G and H are conjugate to each other. Right: The respective values $y(x^*)$ of the neural activity. Note that $y(x_1^*) - y(x_0^*) \geq 1/2$, for all values of the bias.

$\epsilon_w = 0.01$. We constrained the activities of the presynaptic neurons y_j, for consistency, to the interval $[0, 1]$, which is the same interval of postsynaptic firing rates. Generically we considered uni- and bi-modal Gaussian inputs centered around $\bar{y}_j = 0.5$, with individual standard deviations σ_j. We considered in general $\sigma_j = 0.25$ for the direction having the largest variance, the dominant direction, with the other directions having smaller standard deviations, typically by a factor of two.

2.3 EMERGENT SLIDING THRESHOLD

One may invert the sigmoidal $\sigma(x)$ via $x = b - \log\big((1-y)/y\big)$ and express the adaption factors solely in terms of the neural firing rate y. For the Hebbian factor $H(x)$, see Eq. (5), one then finds

$$H(y) = (2y - 1) + 2y(1 - y)\big[b - \log\big((1-y)/y\big)\big] . \tag{15}$$

The bias b hence regulates the crossing point from anti-Hebbian (for low neural activity $y < y_H^*$) to Hebbian learning (for large firing rates $y > y_H^*$), where y_H^* is the root of $H(y)$. y_H^* depends only on b (as shown in Fig. 3), emerging then indirectly from the formulation of the objective function (3), and plays the role of a sliding threshold. This sliding threshold is analogous to the one present in the BCM theory (**Bienenstock et al.**, 1982), which regulates the crossover from anti-Hebbian to Hebbian learning with increasing output activity and which is adapted in order to keep the output activity within a given working regime.

The bias b regulates, in addition to its role determining the effective sliding threshold for synaptic plasticity, the mean firing rate. In principle, one may consider an ad-hoc update rule like $\dot{b} \propto (\mu - \bar{y})$ for the bias, where $\mu = \int y p_\lambda(y) dy$ is some given target firing rate and where $\bar{y}$ would be a sliding average of y. We will however use, alternatively, an information theoretical objective function for the intrinsic adaption of the bias. The Kullback-Leibler divergence

$$D = \int dy\, p(y) \log\left(\frac{p(y)}{p_\lambda(y)}\right), \qquad p_\lambda(y) = \frac{\mathrm{e}^{\lambda y}}{N_\lambda} \tag{16}$$

measures the distance between the actual firing-rate distribution $p(y)$ and a given target distribution $p_\lambda(y)$. It will be minimal if $p_\lambda(y)$ is approximated as well as possible. An exponential target distribution, as selected here, maximizes the information content of the neural activity in terms of information entropy,

given the constraint of a fixed mean μ, both for a finite support $y \in [0, 1]$, as considered here, as well as for an unbounded support, $y > 0$, with N_λ being the appropriate normalization factor. For $\lambda \to 0$ a uniform target distribution is recovered together with $\mu \to 0.5$ and the resulting $p(y)$ becomes symmetric with respect to $y = 0.5$.

Following a derivation which is analogous to the one given above for the case of synaptic flux minimization, one finds Eq. (6) for the adaption rules (**Triesch**, 2007; **Linkerhand and Gros**, 2013b). For the adaption rate ϵ_b for the bias we used in our simulations generically $\epsilon_b = 0.1$, its actual value having only a marginal influence on the overall behavior of the adaption processes.

Minimizing the Kullback-Leibler divergence and the Fisher information are instances of polyhomeostatic optimization (**Marković and Gros**, 2010, 2012), as one targets to optimize an entire probability distribution function, here $p(y)$. An update rule like $\dot{b} \propto (\mu - \bar{y})$ would, on the other side, correspond to a basic homeostatic control, aiming to regulate a single scalar quantity, such as the mean firing rate.

2.4 FIXPOINTS OF THE LIMITING FACTOR

The self-limiting factor $G(x)$ has two roots x_G^*, compare Fig. 3. For $b = 0$ one finds $x_G^* \approx \pm 2.4$ corresponding to firing-rates $y_G^* = 0.083$ and $y_G^* = 0.917$ respectively, compare also Fig. 3. The roots of $G(x)$ are identical with the two minima of the objective function F_{ob}, compare (3). The self-limiting nature of the synaptic adaption rules (4) is a consequence of the two roots of $G(x)$, as larger (in magnitude) membrane potentials will reverse the Hebbian adaption to an anti-Hebbian updating. The roots of $G(x)$ induce, in addition, the tendency of performing a binary classification. As an illustration consider the case of random sequences of discrete input patterns

$$\mathbf{y}^\eta, \qquad \eta = 1, \ldots, N_{patt} , \tag{17}$$

with the number of input patterns N_{patt} being smaller than the number of afferent neurons, $N_{patt} \leq N_w$. The inputs $(y_1, \ldots, y_{N_w}) = \mathbf{y}$ are selected randomly out of the set (17) of N_{patt} patterns and presented consecutively for small time intervals. The synaptic updating rules will then lead, as we have tested through extended simulations, to a synaptic vector $\mathbf{w}$ dividing the space of input patterns into two groups,

$$\begin{aligned} \mathbf{w} \cdot (\mathbf{y}^\eta - \bar{\mathbf{y}}) &= x_G^*(1) & \text{for} & \gamma N_{patt} & \text{states} & \mathbf{y}^\eta \\ \mathbf{w} \cdot (\mathbf{y}^\eta - \bar{\mathbf{y}}) &= x_G^*(2) & \text{for} & (1 - \gamma) N_{patt} & \text{states} & \mathbf{y}^\eta \end{aligned}, \tag{18}$$

which is a solvable set of N_{patt} equations for N_w variables $(w_1, w_2, \ldots)$. Here we have denoted with $x_G^*(1)$ and $x_G^*(2)$ the two distinct roots of $G(x)$, and with $\bar{\mathbf{y}} = \left(\sum_\eta \mathbf{y}^\eta \right) / N_{patt}$ the mean input activity. This outcome of the long-term adaption corresponds to a binary classification of the N_{patt} vectors. The membrane potential $\mathbf{x} = \mathbf{w} \cdot \mathbf{y}$ just takes two values, for all inputs $\mathbf{y}$ drawn from the the set of input patterns.

There is one free parameter in (18), namely the fraction γ and $(1 - \gamma)$ of patterns mapped to $x_G^*(1)$ and $x_G^*(2)$ respectively. This fraction γ is determined self-consistently by the system, through the polyhomeostatic adaption (6) of the bias b, with the system trying to approximate as close as possible the target firing-rate distribution $\propto \exp(\lambda y)$, see Eq. (16).

3 RESULTS

In order to test the behavior of the neuron under rules (4,6) when presented with different input distributions, a series of numerical simulations have been performed. In the following sections, the evolution of the system when faced with static input distributions is first studied. In particular, principal

component extraction and linear discrimination tasks are evaluated. These results are then extended to a scenario of varying input distributions and a fading memory effect is then analyzed.

3.1 PRINCIPAL COMPONENT EXTRACTION

As a first experiment we consider the case of N_w input neurons with Gaussian activity distributions $p(y_j)$. In this setup a single component, namely y_1, has standard deviation σ and all other $N_w - 1$ directions have a smaller standard deviation of $\sigma/2$, as illustrated in Fig. 4(A). We have selected, for convenience, y_1 as the direction of the principal component. The synaptic updating rule (1) is however fully rotational invariant in the space of input activities and the results of the simulations are independent of the actual direction of the principal component. We have verified this independence by running simulations with dominant components selected randomly in the space of input activities.

In Fig. 4 we present the result for $N_w = 100$ afferent neurons and $\lambda = -2.5$ for the target distribution $p_\lambda(y)$, compare Eq. (16) in section 2. The initial synaptic weights $\{w_j\}$ have been randomly drawn from $[-0.005 : 0.005]$ and are hence quite small, such that the learning rule is initially exclusively Hebbian, viz the membrane potential x is substantially smaller than the roots x_G^* of the limiting factor $G(x)$ (compare Fig. 4(B) where x/x_G^* are given by the blue/red dots respectively). Hebbian synaptic growth then eventually leads to larger weights, with the weight along the principal component (here w_1, red line in Fig. 4(F)) becoming very large. At this stage, the membrane potential x starts to cross the roots x_G^* of the limiting factor $G(x)$ and a stationary state results, with the weight along the principal component saturating and with the weights along the non-principal components involved in bounded random drifts. This stationary state, with continuously ongoing online learning, remains stable for arbitrary simulation times.

The firing rate $y(t)$ covers the whole available interval $[0, 1]$, in the stationary state, and a sliding threshold emerges self-consistently. This sliding threshold is given by the root x_H^* of the Hebbian factor $H(x)$; learning is Hebbian/anti-Hebbian for $y > y(x_H^*)$ and $y < y(x_H^*)$ respectively. For our simulation the sliding threshold is about $y(x_H^*) \simeq 0.4$ (green dots in Fig. 4(D)) in the stationary state.

The angle α between the direction of the synaptic weight vector $\mathbf{w}$ and the principal component of input activities is initially large, close to the random value of $\pi/2$, dropping close to zero with forthgoing synaptic adaption, as shown in Fig. 4(C), a consequence of the growth of w_1. In Fig. 4(E) we plot the distribution of the w_j, with a separate scale for the principal component, here $w_1 \approx 9.1$ (as averaged over 100 runs). The small components are Gaussian distributed around zero with a standard deviation of $\sigma_w^{(non)} \approx 0.23$, we have hence a large signal-to-noise ratio of $S_w = |w_1|/\sigma_w^{(non)} \approx 9.1/0.23 \approx 40$.

We also present in the inset of Fig. 4(C) a comparison between the actual firing-rate distribution $p(y)$ in the stationary state and the exponential target distribution $\propto \exp(\lambda y)$, entering the Kullback-Leibler divergence, see Eq. (16).

3.2 SIGNAL-TO-NOISE SCALING

For synaptic adaption rules to be biologically significant they should show stable performance even for large numbers N_w of afferent neurons, without the need for fine-tuning of the parameters. This is the case for our plasticity rules.

In Fig. 5 we present the scaling behavior of the synaptic weight configuration. We consider both a large range for the number N_w of afferent neurons and an extended range for the incoming signal-to-noise ratio. The input activity distributions $p(y_j)$ are Gaussians with standard deviations $\sigma_j = \sigma_\perp$ for $(j = 2, \ldots, N_w)$, and with the dominant direction having a width σ_1. We define the incoming signal-to-noise ratio as $S_i = \sigma_1/\sigma_\perp$, and investigate values for S_i of 2:1, 4:1, 8:1, 16:1 and 32:1. Shown in Fig. 5 is the evolution of the outgoing signal-to-noise ratio, as a function of inputs N_w, and the evolution of the angle α. All simulation parameters are kept otherwise constant.
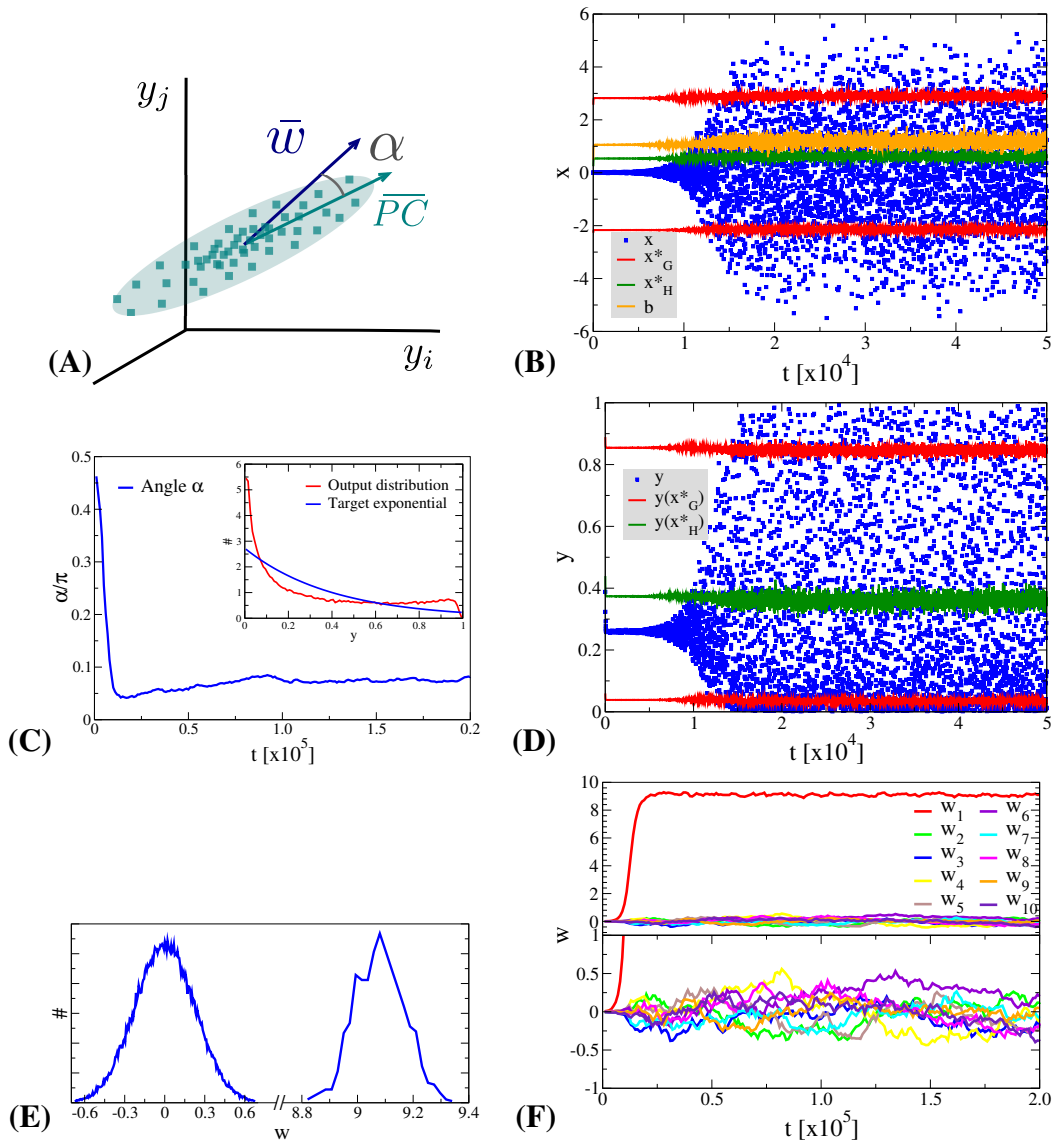
Figure 4. Alignment to the principal component. Simulation results for a neuron with $N_w = 100$ input neurons with Gaussian input distributions with one direction (the principal component) having twice the standard deviation than the other $N_w - 1$ directions. **(A)** Illustration of the input distribution density $p(y_1, y_2, \dots)$, with the angle α between the direction of the principal component ($\overline{PC}$) and $\overline{w}$, the synaptic weight vector. **(B)** Time series of the membrane potential x (blue), the bias b (yellow), the roots x_G^* of the limiting factor $G(x)$ (red) and the root x_H^* of the Hebbian factor $H(x)$ (green). **(C)** The evolution of the angle α of the synaptic weight vector $\mathbf{w}$ with respect to the principal component and (inset) the output distribution $p(y)$ (red) with respect to the target exponential (blue). **(D)** Time series of the output y (blue) and of the roots y_G^* of the limiting factor $G(y)$ (red) and the root y_H^* of the Hebbian factor $H(y)$ (green). **(E)** Distribution of synaptic weights $p(w)$ in the stationary state for large times. **(F)** Time evolution of the first ten synaptic weights $\{w_j\}$, separately for the principal component (upper panel) and for nine other orthogonal directions (lower panel).

We define the outgoing signal-to-noise ratio as $S_w = |w_1|/\sigma_w^{(non)}$ where w_1 is the synaptic weight along the principal component and $\sigma_w^{(non)}$ the standard deviation of the remaining synaptic weights (compare Eq. (22) of the appendix). The outgoing signal-to-noise ratio is remarkably independent of the actual number N_w of afferent neurons. S_w shows, in addition, a threshold behavior, remaining finite even for data input streams characterized by small S_i. For large value of incoming signal-to-noise ratio a linear scaling $S_w \propto S_i$ is recovered.
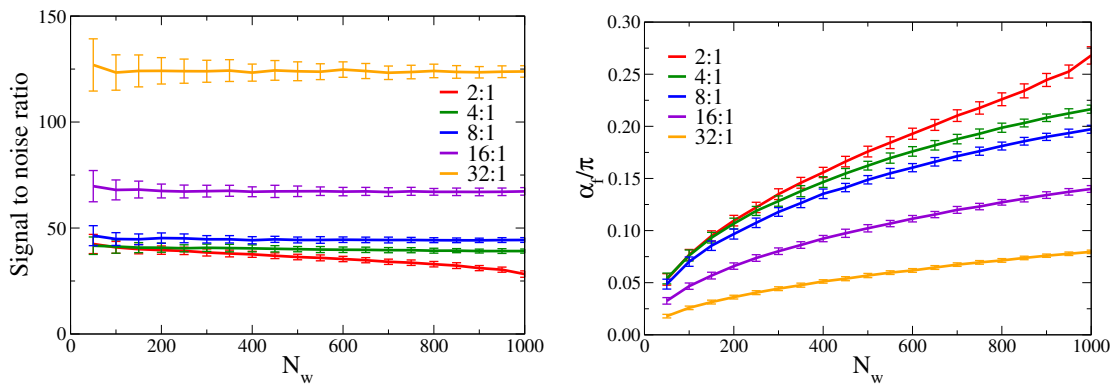
Figure 5. Scaling of the adaption rules with the number of afferent neurons. For constant simulating parameters the signal to noise ratio (left), defined as the ratio $|w_1|/\sigma_w^{(non)}$, where w_1 is the synaptic strength parallel to the principal component and $\sigma_w^{(non)}$ the standard deviation of the orthogonal synaptic directions, compare Eq. (22), and the mean angle (right), of the synaptic weight vector with respect to the principal component. Shown are results for a range, 2:1, 4:1, 8:1, 16:1 and 32:1, of the incoming signal-to-noise ratios, defined as the ratio of the standard deviations between the large and the small components of the distributions of input activities $p(y_j)$. The outgoing signal-to-noise ratio $|w_1|/\sigma_w^{(non)}$ remains essentially flat, as a function of N_w; the increase observed for the average angle α is predominantly a statistical effect, caused by the presence of an increasingly large number of orthogonal synaptic weights. The orthogonal weights are all individually small in magnitude, but their statistical influence sums up increasingly with raising N_w.

Regarding the angle α, the performance deteriorates, which increases steadily with N_w. This is however a dominantly statistical effect. In the appendix we show how the angle α increases with N_w for a constant outgoing signal-to-noise ratio S_w. This effect is then just a property of angles in large dimensional spaces and is independent of the learning rule employed.

It is interesting to compare the simulation results with other updating rules, like Oja's rule (**Oja**, 1997),

$$\dot{w}_j \;=\; \epsilon_{oja} \left[y(y_j - \bar{y}_j) - \alpha\, y^2 w_j \right] \;. \tag{19}$$

The original formulation used $\alpha = 1$ for the relative weighting of the decay term in (19). We find however, for the case of non-linear neurons considered here, that Oja's rule does not converge for $\alpha \gtrsim 0.1$. For the results presented in Fig. 6 we adapted the bias using (6) both when using Oja's rule (19) and for our plasticity rule (4). The parameter ϵ_{oja} was chosen such that the learning times (or the number of input patterns) needed for convergence matched, in this case $\epsilon_{oja} = 0.1$. With Oja's rule, arbitrarily large outgoing signal-to-noise ratios are achievable for $\alpha \to 0$. In this case the resulting $p(y)$ becomes binary, as expected. There is hence a trade-off and only intermediate values for the outgoing signal-to-noise ratio are achievable for smooth firing-rate distributions $p(y)$. Note that Sanger's rule (**Sanger**, 1989) reduces to Oja's rule for the case of a single neuron, as considered here.

We also attempted to compare with the results of the BCM theory (**Bienenstock et al.**, 1982; **Intrator and Cooper**, 1992; **Cooper and Bear**, 2012). The BCM update rule also finds nicely the direction of the principal component, but runaway synaptic growth occurs generically in the case of the type of neurons considered in our study, being non-linear and having an maximal possible firing rate, with $y \in [0, 1]$. This is due to the fact that the upper cut-off of the firing rate preempts, in general, the sliding threshold to raise to values necessary to induce a large enough amount of synaptic weight decay. For the input distributions used throughout this study we could not avoid runaway synaptic growth for the BCM rule.

3.3 LINEAR DISCRIMINATION

An important question regards the behavior of neural learning rules when no distinct principal component is present in the data input stream. In Fig. 7 we present data for the situation where two dominant directions have the same standard deviation $\sigma \approx 0.22$, here for $p(y_1)$ and $p(y_2)$, with the remaining $N_w - 2$ directions
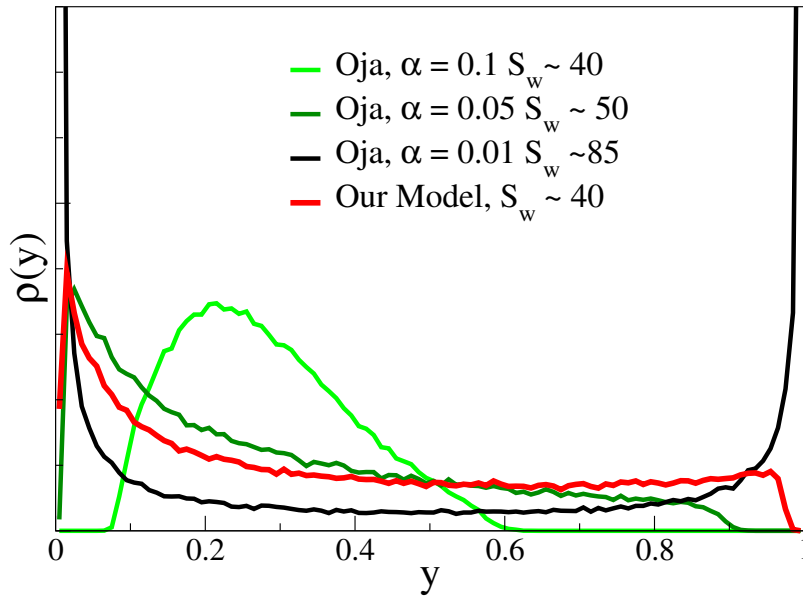
Figure 6. The output distribution function and the signal to noise ratio. The averaged firing-rate distributions $p(y)$ for $N_w = 100$ and and the parameter set used previously, compare Fig. 5. In comparison the $p(y)$ resulting when using a modified Oja's rule, see Eq. (19), for the synaptic plasticity. Depending on the parameter ϵ, controlling the strength of the weight decay in (19), arbitrary large signal-to-noise ratios S_w can be achieved, on the expense of obtaining binary output distributions. Note that the form of $p(y)$ is roughly comparable, for similar signal-to-noise ratios, for the two approaches. The output tends to cluster, however, around the target mean for smaller S_w and Oja's rule.

having a smaller standard deviation $\sigma/4$. In our experiment the first direction, y_1 is a unimodal Gaussian, as illustrated in Fig. 7(A), with the second direction, y_2 being bimodal. The two superposed Gaussian distributions along y_2 have individual widths $\sigma/4$ and the distance between the two maxima has been adjusted so that the overall standard deviation along y_2 is also σ.

The synaptic weight vector aligns, for most randomly drawn starting ensembles $\{w_j\}$, with the bimodal direction, as shown in Fig. 7(C). In this case the system tries to adjust its parameters, namely the synaptic weights and the bias b so that the two peaks of the bimodal principal component are close to the two zeros x_G^* (red symbols in Fig. 7(B)) of the limiting factor $G(x)$ in the adaption rule (4). This effect is clearly present in the results for the membrane potential (blue symbols in Fig. 7(B)), clustering around the roots of $G(x)$. The system performs, as a result, a linear discrimination with a bimodal output firing rate, presented in Fig. 7(D).

One possibility to characterize the deviation of a probability distribution from a Gaussian is the excess kurtosis κ (**DeCarlo**, 1997),

$$\kappa = \frac{Q_j}{\sigma_j^4} - 3, \qquad Q_j = \int (y_j - \bar{y}_j)^4 p(y_j) dy_j, \qquad \sigma_j^2 = \int (y_j - \bar{y}_j)^2 p(y_j) dy_j , \qquad (20)$$

with the normal distribution having, by construction, a vanishing $\kappa \to 0$. The excess kurtosis tends to be small or negative on a finite support $p_j \in [0, 1]$. Distributions characterized by a positive κ show pronounced tails. This statement also holds for truncated Gaussians, as used in our simulations. We have generalized the experiment presented in Fig. 7 by studying the pairwise competition between three distributions having all the same standard deviation σ, but varying values of κ, compare Fig. 7(E): A bimodal distribution with $\kappa = -1.69$, a unimodal Gaussian with $\kappa = -0.63$ and a unimodal double exponential with $\kappa = -0.43$.
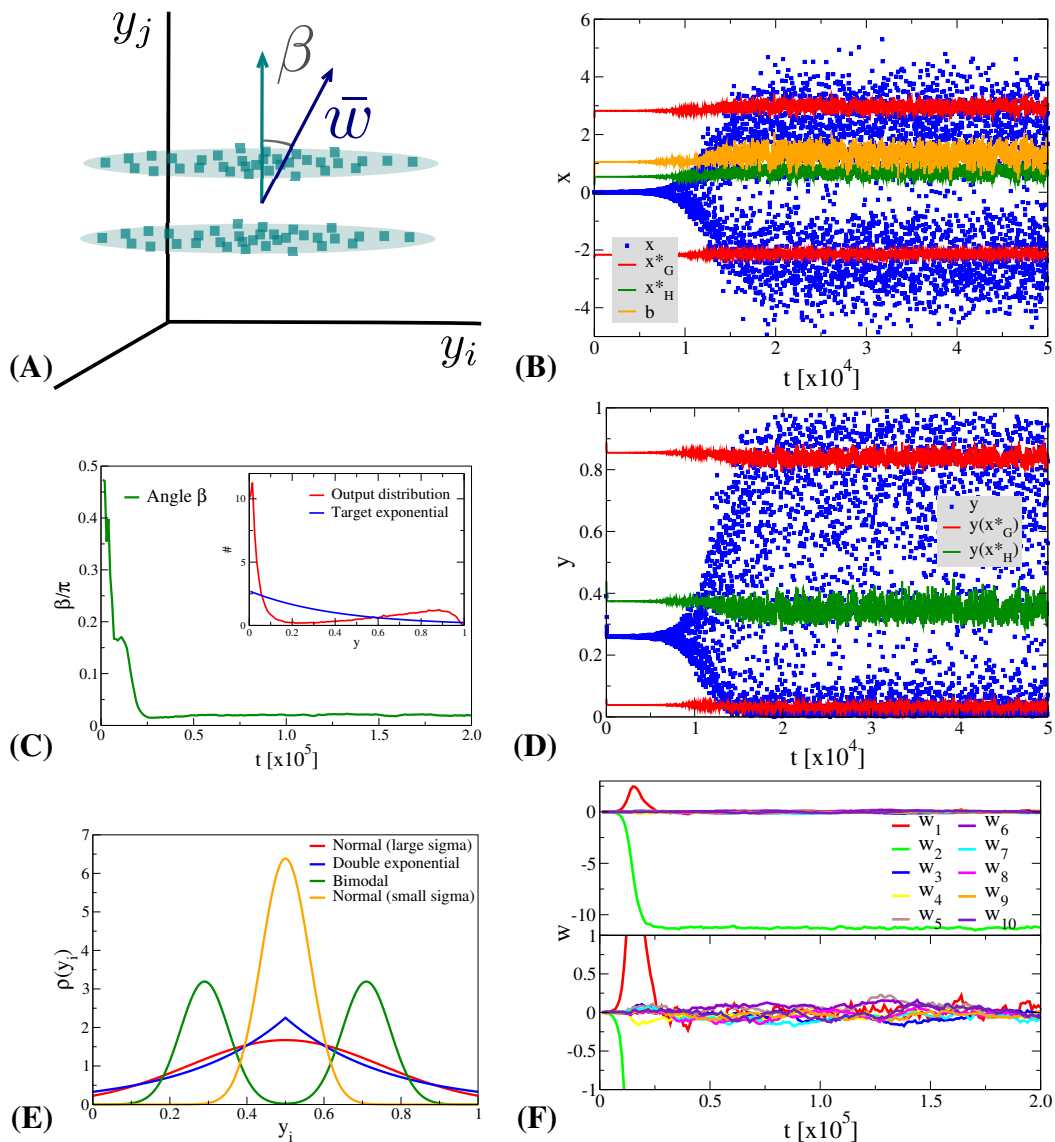
Figure 7. Linear discrimination of bimodal input distributions. Simulation results for a neuron with $N_w = 100$ with two directions having the same variance (but one being bimodal) and the other $N_w - 2$ directions having a standard deviation four times smaller. **(A)** Illustration of the input distribution density $p(y_1, y_2, \ldots)$. **(B)** Time series of the membrane potential x (blue), the bias b (yellow), the roots x_G^* of the limiting factor $G(x)$ (red) and the root x_H^* of the Hebbian factor $H(x)$ (green). **(C)** The evolution of the angle β of the synaptic weight vector $\mathbf{w}$ with respect to the axis linking the two ellipsoids and (inset) the output distribution $p(y)$ (red) in comparison to the target exponential (blue). **(D)** Time series of the output y (blue) and of the roots y_G^* of the limiting factor $G(y)$ (red) and the root y_H^* of the Hebbian factor $H(y)$ (green). **(E)** Illustration of the distribution functions used, the bimodal competing with the Normal distributed (alternatively with a double exponential) having the same variance, all other directions being normally distributed with a four times smaller standard distribution. **(F)** Time evolution of the first ten synaptic weights $\{w_j\}$, separately for the principal component (upper panel) and for nine other orthogonal directions (lower panel).

Running the simulation one thousand times, with randomly drawn initial conditions, the direction with lower κ was selected 88.8% / 65.4% / 64.0% of the times when the competing directions were bimodal vs. double exponential / Gaussian vs. double exponential / bimodal vs. Gaussian. In none of the cases would both the first and the second synaptic weights, w_1 and w_2, acquire large absolute values.

The underlying rationale for the updating rules favoring directions with negative excess kurtosis can be traced back to the inherent symmetry $F_{ob}(-x, 1 - y)) = F_{ob}(x, y))$ of the objective function (3), which

in turn is a consequence of treating both large and small firing rates on an equal footing in F_{ob}. There are two equivalent minima for F_{ob} to which the maxima of a binary distribution are mapped, as discussed in section 2.

We have repeated this simulation using the modified Oja's rule (19), using $\alpha = 0.1$ and $\epsilon_{oja} = 0.1$. We find a very distinct sensitivity, with the relative probability for a certain input direction to be selected being 97.0% / 99.8% / 42.1% when the competing directions were bimodal vs. double exponential / Gaussian vs. double exponential / bimodal vs. Gaussian. Note that all our input distributions are centered around 0.5 and truncated to $[0, 1]$. Oja's rule has a preference for unimodal distributions and a strong dislike of double exponentials. However, the excess kurtosis does not seem to be a determining parameter, within Oja's rule, for the directional selectivity.

3.4 CONTINUOUS ONLINE LEARNING - FADING MEMORY

Another aspect of relevance concerns the behavior of synaptic plasticity rules for continuous online learning. A basic requirement is the absence of runaway growth effects in the presence of stationary input statistics. But how should a neuron react when the statistics of the afferent input stream changes at a certain point? Should it adapt immediately, at a very short time scale or should it show a certain resilience, adapting to the new stimuli only when these show a certain persistence?

We have examined the behavior of the adaption rules upon a sudden change of firing-rate statistics of the afferent neurons. We find, as presented in Fig. 8, that the new statistics is recognized autonomously, with a considerable resilience to unlearn the previously acquired information about the statistics of the input data stream. The synaptic plasticity rules (4) do hence incorporate a fading memory.

In our experiment we considered $N_w = 100$ afferent neurons, with Gaussian firing distributions having standard deviation σ for the principal component and $\sigma/2$ for the remaining $N_w - 1$ directions. The sign of the synaptic weights are not of relevance, as the input distributions $p(y_j)$ are symmetric with respect to their means, taken to be 0.5. The direction of the principal component is then changed several times, everything else remaining otherwise unchanged.

The starting configuration $\{w_j\}$ of synaptic weights has been drawn randomly from $[-0.005 : 0005]$ and the initial learning is fast, occurring on a time scale of $T_{initial} \approx 10^4$ updatings, compare Fig. 4, using the same updating rates $\epsilon_w = 0.01$ and $\epsilon_b = 0.1$ as throughout this paper. The time the neuron takes to adapt to the new input statistics is however of the order of $T_{unlearn} \approx 10^6$, viz about two orders of magnitude larger than $T_{initial}$. New information is hence acquired at a slower rate; the system shows a substantial resilience to unlearn previously acquired memories.

One can observe in Fig. 8 an overshoot of the principal synaptic weight, just before the unlearning starts, as the system tries to keep the membrane potential x within its working regime, compare Fig. 4(D). The system reacts by increasing the largest synaptic weight when the variance of the input drops along the corresponding afferent direction, before it can notice that the principal component of the afferent activities has also changed.

Also included in the simulation presented in Fig. 8 is a phase without any principal component, the statistics of all incoming $p(y_j)$ being identical, viz with the covariance matrix being proportional to unity. One notices that the neuron shows a marked resilience to forget the previously acquired knowledge, taking about $5 \cdot 10^7$ updates in order to return to a fully randomized drifting configuration of synaptic weights $\{w_j\}$. The synaptic plasticity rule (4) hence leads to an extended fading memory, which we believe to be a consequence of its multiplicative structure.

For comparison we have repeated the same experiment using the modified Oja's rule (19), using $\alpha = 0.1$ (which yields the same signal-to-noise ratio, compare Fig. 6), and $\epsilon_{oja} = 0.1$, such that the initial learning rates (achieving 90% of the stationary value for the principal component) are comparable for both updating
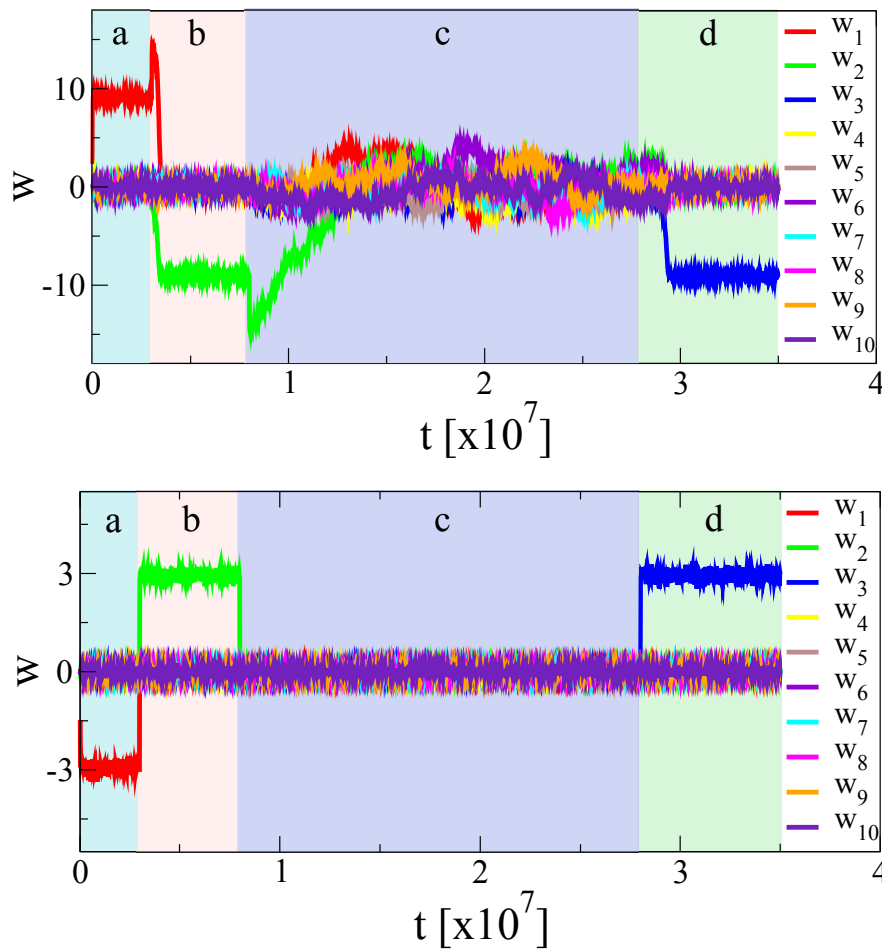
Figure 8. Continuous online learning and weak forgetting. The effect of changing the statistics of the input firing rates $p(y_j)$. During (a), (b) and (d) the principal axis is along y_1, y_2 and y_3 respectively, during (c) there in no principal component. There are $N_w = 100$ afferent neurons, shown is the time evolution of the first ten synaptic weights. The standard deviations of the afferent neurons is σ for the principal direction, if there is any, and $\sigma/2$ for all other directions, compare Fig. 4. During(c) all inputs have identical standard deviations $\sigma/2$. The initial weight distribution is randomly drawn. The top/bottom panel show the results using respectively our synaptic updating rule (4) and Oja's rule (19). Note that it takes considerably longer, for our updating rule, to unlearn than to learn from scratch. Learning and unlearning occurs, on the other side, at the same timescale for Oja's rule.

rules. We also kept the same updating (6) for the bias. For Oja's rule learning and unlearning occurs on very similar time scales, reacting immediately to changes in the statistics of the input activities.

It is presently not entirely clear which form of unlearning is present in the brain, on the level of individual neurons. While studies in prefrontal cortex have shown full learning and unlearning of different categories in binary classification tasks, related in this context to the concept of adaptive coding (**Duncan**, 2001), more complex behavioral responses tend however to exhibit slow or incomplete unlearning such as extinction of paired cue - response associations, in the context of Pavlovian conditioning (**Myers and Davis**, 2002; **Quirk and Mueller**, 2007). It is also conceivable that a fading memory may possibly be advantageous in the context of noisy environments with fluctuating activity statistics.

4 DISCUSSION

Objective functions based on information theoretical principles play an important role in neuroscience (**Intrator and Cooper**, 1992; **Lengellé and Denoeux**, 1996; **Goodhill and Sejnowski**, 1997; **Kay and Phillips**, 2011) and cognitive robotics (**Sporns and Lungarella**, 2006; **Ay et al.**, 2008). Many objective functions investigated hitherto use either Shannon's information directly, or indirectly by considering related measures, like predictive and mutual entropy (**Kraskov et al.**, 2004), or the Kullback-Leibler divergence. Objective functions are instances, from a somewhat larger perspective, of generating functionals, as they are normally used to derive equations of motion for the neural activity, or to deduce adaption rules for secondary variables like synaptic weights or intrinsic parameters. Here we discuss an objective function which may be either motivated by its own virtue, as discussed in section 2, or by considering the Fisher information as a generating functional.

The Fisher information encodes the sensitivity of a given probability distribution function, in our case the distribution of neural firing rates, with respect to a certain parameter of interest. Cognitive information processing in the brain is all about changing the neural firing statistics and we hence believe that the Fisher information constitutes an interesting starting point from where to formulate guiding principles for plasticity in the brain or in artificial systems. In particular, we have examined the Fisher information with respect to changes of the synaptic weights. Minimizing this objective function, which we denoted as the synaptic flux, we find self-limiting adaption rules for unsupervised and autonomous learning. The adaption rules are Hebbian, with the self limitation leading to synaptic competition and an alignment of the synaptic weight vector with the principal component of the input data stream.

Synaptic plasticity rules for rate encoding neurons are crucial for artificial neural networks used for cognitive tasks and machine learning, and important for the interpretation of the time-averaged behavior of spiking neurons. In this context our adaption rules make two predictions, which one may eventually test experimentally. The first prediction concerns the adaption in the situation where more than one dominant component is present in the space of input activities. Our model implies for this case a robust tendency for the synaptic weight vector to favor directions in the space of input activities being bimodal, characterized by a negative kurtosis.

Our adaption rules have a second implication, regarding the robustness of acquired memories with respect to persistent changes of the statistics of the input activities, in the context of continuous and unsupervised online learning. We predict that it is considerably easier for the neuron to detect relevant features in the space of input activities when starting from a virgin state of a random synaptic configuration. New features will still be extracted from the stream of input activities, and old ones unlearned at the same time, once the initial synaptic adaption process has been completed, albeit at a much slower pace. This feature can be interpreted as a sturdy fading memory.

We have extensively examined the robustness of the behavior of the synaptic plasticity rules upon variation of the simulation setup. All results presented here remain fully valid when changing, e.g. the adaption rate ϵ_w, in particular we have examined $\epsilon_w = 0.1$ and $\epsilon_w = 0.001$. We have also studied other forms of input activities $p(y_j)$ and found only quantitative changes for the response. For example, we have considered exponentially distributed input statistics, as a consistency check with the target output distribution function. We hence believe that the here proposed synaptic plasticity rules are robust to a considerable degree, a prerequisite for viable plasticity rules, both in the context of biological and artificial systems.

The synaptic plasticity rule (4) is a product of two conjugate factors, the limiting factor $G(x)$ and the Hebbian factor $H(x)$. Runaway synaptic growth occurs, as we have verified numerically, when setting $G(x)$ to a constant. Unlimited synaptic growth occurs despite the emergence of a sliding threshold (see Eq. (15) of section 2) as the firing rate $y(t) \in [0,1]$ is bounded. Runaway synaptic growth results in increasing (positive and negative) large membrane potentials $x(t)$, with the firing becoming binary, accumulating at the boundaries, viz $y \to 0$ and $y \to 1$.

Finally we comment on the conceptual foundations of this work. The adaptive time evolution of neural networks and the continuous reconfiguration of synaptic weights may be viewed as a self-organizing processes guided by certain target objectives (**Gros**, 2010; **Prokopenko**, 2009; **Friston**, 2010; **Linkerhand and Gros**, 2013a). A single objective function will in general not be enough for generating dynamics of sufficient complexity, as necessary for neural circuitry or synaptic reconfiguration processes. It has indeed been noted that the interplay between two or more generating functionals may give rise to highly nontrivial dynamical states (**Linkerhand and Gros**, 2013a; **Gros**, 2014).

In this context, it is important to note that several generating functionals may in general not be combined to a single overarching objective function. Dynamical systems can hence show, under the influence of competing objective functions, complex self-organizing behavior (**Linkerhand and Gros**, 2013a; **Gros**, 2014). In the present work we propose that the interplay between two specific objective functions, namely the Fisher information for the synaptic flux and the Kullback-Leibler divergence for the information content of the neural firing rate, give rise, quite naturally, to a set of viable adaption rules for self-limiting synaptic and intrinsic plasticity rules.

5 ACKNOWLEDGMENTS

6 APPENDIX: MODELING ADAPTION FOR LARGE NUMBERS OF TRANSVERSAL DIRECTIONS

For simulations with N_w Gaussian input distributions $p(y_j)$ the synaptic weight vector adapts to

$$\mathbf{w} = (w_1, w_2, \ldots, w_{N_w}), \qquad w_1 \gg w_k \qquad (k \geq 2), \qquad (21)$$

when $p(y_1)$ is assumed to have the largest standard deviation σ_1, with all other $p(y_k)$, for $k = 2, \ldots, N_w$ having a smaller standard deviation σ_k. The angle α between the synaptic weight vector and the direction $(1, 0, \ldots, 0)$ of the principal component is hence given by

$$\cos(\alpha) = \frac{w_1}{\sqrt{w_1^2 + \sum_{k>1} w_k^2}} = \frac{w_1}{\sqrt{w_1^2 + (N_w - 1)\left(\sigma_w^{(non)}\right)^2}} \approx \frac{1}{\sqrt{N_w}} \frac{w_1}{\sigma_w^{(non)}}, \qquad (22)$$

where we have defined with $\sigma_w^{(non)} = \left(\sum_{k>1} w_k^2\right)/(N_w - 1)$ the averaged standard deviation of the non-principal components (which have generically a vanishing mean). In our simulation we find, compare Fig. 5, an outgoing signal-to-noise ratio $S_w = |w_1|/\sigma_w^{(non)}$ which is remarkably independent of N_w and hence that α approaches $\pi/2$ like $\pi/2 - r/\sqrt{N_w}$ in the limit of large numbers $N_w \to \infty$ of afferent neurons, where r is a constant, independent of N_w. This statistical degradation of the performance, in terms of the angle α, is hence a variant of the well known curse of dimensionality (**Jain et al.**, 2000).

REFERENCES

Abbott, L. F. and Nelson, S. B. (2000), Synaptic plasticity: taming the beast, *Nature Neuroscience*, 3, 1178–1183

Ay, N., Bertschinger, N., Der, R., Güttler, F., and Olbrich, E. (2008), Predictive information and explorative behavior of autonomous robots, *The European Physical Journal B*, 63, 3, 329–339

Bell, A. J. and Sejnowski, T. J. (1995), An information-maximization approach to blind separation and blind deconvolution, *Neural Computation*, 7, 6, 1129–1159

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982), Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex, *The Journal of Neuroscience*, 2, 1, 32–48

Brunel, N. and Nadal, J.-P. (1998), Mutual information, fisher information, and population coding, *Neural Computation*, 10, 7, 1731–1757

Chagnac-Amitai, Y., Luhmann, H. J., and Prince, D. A. (1990), Burst generating and regular spiking layer 5 pyramidal neurons of rat neocortex have different morphological features, *Journal of Comparative Neurology*, 296, 4, 598–613

Cooper, L. N. and Bear, M. F. (2012), The bcm theory of synapse modification at 30: interaction of theory with experiment, *Nature Reviews Neuroscience*, 13, 11, 798–810

DeCarlo, L. T. (1997), On the meaning and use of kurtosis, *Psychological Methods*, 2, 3, 292–307

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012), How does the brain solve visual object recognition?, *Neuron*, 73, 3, 415–434

Dong, D. W. and Hopfield, J. J. (1992), Dynamic properties of neural networks with adapting synapses, *Network: Computation in Neural Systems*, 3, 3, 267–283

Duncan, J. (2001), An adaptive coding model of neural function in prefrontal cortex, *Nature Reviews Neuroscience*, 2, 11, 820–829

Elliott, T. (2003), An analysis of synaptic normalization in a general class of hebbian models, *Neural Computation*, 15, 4, 937–963

Feldman, D. E. (2012), The spike-timing dependence of plasticity, *Neuron*, 75, 4, 556–571

Friston, K. (2010), The free-energy principle: a unified brain theory?, *Nature Reviews Neuroscience*, 11, 2, 127–138

Goodhill, G. J. and Barrow, H. G. (1994), The role of weight normalization in competitive learning, *Neural Computation*, 6, 2, 255–269

Goodhill, G. J. and Sejnowski, T. J. (1997), A unifying objective function for topographic mappings, *Neural Computation*, 9, 6, 1291–1303

Gros, C. (2010), Complex and adaptive dynamical systems: A primer (Springer Verlag)

Gros, C. (2014), Generating functionals for guided self-organization, in M. Prokopenko, ed., Guided Self-Organization: Inception (Springer), 53–66

Gutnisky, D. A. and Dragoi, V. (2008), Adaptive coding of visual information in neural populations, *Nature*, 452, 7184, 220–224

Hebb, D. O. (2002), The organization of behavior: A neuropsychological theory (Psychology Press)

Huber, P. J. (1985), Projection pursuit, *The annals of Statistics*, 435–475

Intrator, N. and Cooper, L. N. (1992), Objective function formulation of the bcm theory of visual cortical plasticity: Statistical connections, stability conditions, *Neural Networks*, 5, 1, 3–17

Jain, A. K., Duin, R. P. W., and Mao, J. (2000), Statistical pattern recognition: A review, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22, 1, 4–37

Kay, J. W. and Phillips, W. (2011), Coherent infomax as a computational goal for neural systems, *Bulletin of mathematical biology*, 73, 2, 344–372

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004), Estimating mutual information, *Physical review E*, 69, 6, 066138

Lengellé, R. and Denoeux, T. (1996), Training mlps layer by layer using an objective function for internal representations, *Neural Networks*, 9, 1, 83–97

Linkerhand, M. and Gros, C. (2013a), Generating functionals for autonomous latching dynamics in attractor relict networks, *Scientific Reports (in press)*

Linkerhand, M. and Gros, C. (2013b), Self-organized stochastic tipping in slow-fast dynamical systems, *Mathematics and Mechanics of Complex Systems*, 1-2, 129

Lisman, J. and Spruston, N. (2010), Questions about stdp as a general model of synaptic plasticity, *Frontiers in Synaptic Neuroscience*, 2

Lisman, J. E. (1997), Bursts as a unit of neural information: making unreliable synapses reliable, *Trends in neurosciences*, 20, 1, 38–43

Marković, D. and Gros, C. (2010), Self-organized chaos through polyhomeostatic optimization, *Physical Review Letters*, 105, 6, 068702

Marković, D. and Gros, C. (2012), Intrinsic adaptation in autonomous recurrent neural networks, *Neural Computation*, 24, 2, 523–540

Miller, K. D. and MacKay, D. J. (1994), The role of constraints in hebbian learning, *Neural Computation*, 6, 1, 100–126

Myers, K. M. and Davis, M. (2002), Behavioral and neural analysis of extinction, *Neuron*, 36, 4, 567–584

Nagy, A. (2003), Fisher information in density functional theory, *The Journal of chemical physics*, 119, 18, 9401–9405

Oja, E. (1992), Principal components, minor components, and linear neural networks, *Neural Networks*, 5, 6, 927–935

Oja, E. (1997), The nonlinear pca learning rule in independent component analysis, *Neurocomputing*, 17, 1, 25–45

Paradiso, M. (1988), A theory for the use of visual orientation information which exploits the columnar structure of striate cortex, *Biological cybernetics*, 58, 1, 35–49

Prokopenko, M. (2009), Guided self-organization, *HFSP Journal*, 3, 287–289

Prokopenko, M., Lizier, J. T., Obst, O., and Wang, X. R. (2011), Relating fisher information to order parameters, *Physical Review E*, 84, 4, 041116

Quirk, G. J. and Mueller, D. (2007), Neural mechanisms of extinction learning and retrieval, *Neuropsychopharmacology*, 33, 1, 56–72

Reginatto, M. (1998), Derivation of the equations of nonrelativistic quantum mechanics using the principle of minimum fisher information, *Physical Review A*, 58, 1775–1778

Sanger, T. D. (1989), Optimal unsupervised learning in a single-layer linear feedforward neural network, *Neural networks*, 2, 6, 459–473

Seung, H. and Sompolinsky, H. (1993), Simple models for reading neuronal population codes, *Proceedings of the National Academy of Sciences*, 90, 22, 10749–10753

Shouval, H. Z., Wang, S. S.-H., and Wittenberg, G. M. (2010), Spike timing dependent plasticity: a consequence of more fundamental learning rules, *Frontiers in Computational Neuroscience*, 4

Simoncelli, E. P. and Olshausen, B. A. (2001), Natural image statistics and neural representation, *Annual review of neuroscience*, 24, 1, 1193–1216

Sinz, F. and Bethge, M. (2013), Temporal adaptation enhances efficient contrast gain control on natural images, *PLoS computational biology*, 9, 1, e1002889

Sporns, O. and Lungarella, M. (2006), Evolving coordinated behavior by maximizing information structure, in Artificial life X: proceedings of the tenth international conference on the simulation and synthesis of living systems, 323–329

Triesch, J. (2007), Synergies between intrinsic and synaptic plasticity mechanisms, *Neural Computation*, 19, 4, 885–909

Turrigiano, G. G. and Nelson, S. B. (2000), Hebb and homeostasis in neuronal plasticity, *Current opinion in neurobiology*, 10, 3, 358–364

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011), Transfer entropya model-free measure of effective connectivity for the neurosciences, *Journal of computational neuroscience*, 30, 1, 45–67

Wiskott, L. and Sejnowski, T. J. (2002), Slow feature analysis: Unsupervised learning of invariances, *Neural Computation*, 14, 4, 715–770