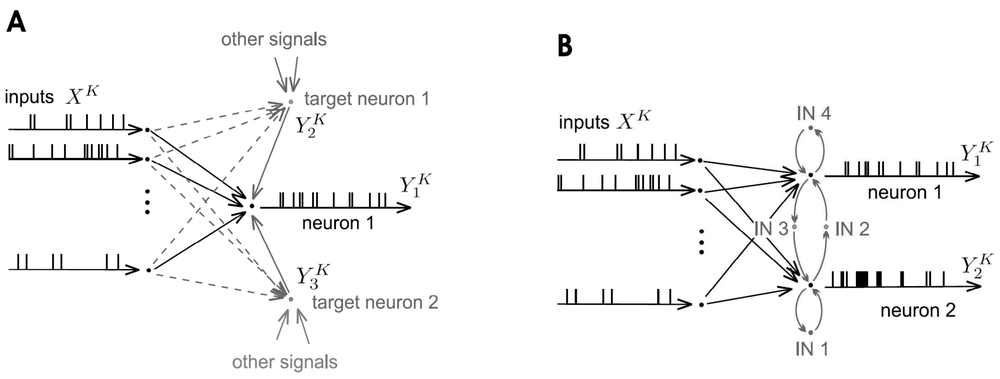
Information Bottleneck Optimization and Independent Component Extraction with Spiking Neurons

Stefan Klampfl, Robert A. Legenstein, Wolfgang Maass

1 Introduction

We show that information bottleneck optimization (IB) [1] and extraction of independent components (IC) [2] can be implemented by stochastic spiking neurons with refractoriness. Both have attracted substantial interest as general principles for unsupervised learning in biological systems, however, concrete learning rules that implement these principles for spiking neurons have still been missing.

Learning situations:



- **Information Bottleneck (IB)** (Fig. A): We want to maximize the mutual information between the output Y_1^K of a learning neuron and the activity of several target signals $Y_2^K, Y_3^K, \ldots$ (which can be functions of the inputs plus some other external signals) while at the same time keeping the mutual information between the inputs X^K and the output Y_1^K as low as possible.
- **Independent Component (IC) Extraction** (Fig. B): We want two neurons that receive the same inputs X^K at their synapses to maximize their information transmission and simultaneously, with the help of interneurons, keep their outputs Y_1^K and Y_2^K statistically independent.

In both learning situations we additionally want to keep the output of the learning neurons within a biologically realistic range by maintaining a constant output firing rate. We extend the approach in [3] where it has been shown that maximizing information transmission for a spiking neuron yields a generalized BCM rule [5].

2 Neuron Model

We use the model from [3], a stochastically spiking neuron with refractoriness, where the membrane potential of neuron i at time $t^k = k\Delta t$ is given as the sum over all postsynaptic potentials at synapses $j = 1, \ldots, N$:

$$u_i(t^k) = u_r + \sum_{j=1}^{N} \sum_{n=1}^{k} w_{ij} \epsilon(t^k - t^n) x_j^n, \tag{1}$$

where $u_r = -70\text{mV}$ is the resting potential and w_{ij} is the weight of synapse j. $x_j^n \in \{0, 1\}$ denotes the presence of a input spike at synapse j at time t^n, which evokes a postsynaptic potential (PSP) with time course $\epsilon(t^k - t^n)$.
At each time step the neuron fires with a certain probability that depends on the current membrane potential and refractory state. This neuron model is a stochastic version of the integrate-and-fire model [7]. The probability of firing for neuron i in the k-th time step is given by

$$\rho_i^k = 1 - \exp[-g(u_i(t^k))R_i(t^k)\Delta t] \approx g(u_i(t^k))R_i(t^k)\Delta t, \tag{2}$$

where $g(u)$ is a smooth increasing function of the membrane potential u and $R(t) \in [0; 1]$ is the refractory variable.

3 Learning Rules

Consider spike trains X^K, Y_1^K, and Y_2^K of length $K\Delta t$. The objective functions to be maximized for the IB and IC case are given as

$$L^{IB} = -I(\mathbf{X}^K; \mathbf{Y}_1^K) + \beta I(\mathbf{Y}_1^K; \mathbf{Y}_2^K) - \gamma D_{KL}(P(Y_1^K) || \tilde{P}(Y_1^K)), \tag{3}$$

and

$$L_i^{IC} = I(\mathbf{X}^K; \mathbf{Y}_i^K) - \beta I(\mathbf{Y}_1^K; \mathbf{Y}_2^K) - \gamma D_{KL}(P(Y_i^K) || \tilde{P}(Y_i^K)), \tag{4}$$

where

$I(\mathbf{X}^K; \mathbf{Y}_i^K)$	mutual information between input spike trains X^K and output spike train Y_i^K of neuron i
$I(\mathbf{Y}_1^K; \mathbf{Y}_2^K)$	mutual information between spike trains Y_1^K and Y_2^K
$D_{KL}(P(Y_i^K) \|\| \tilde{P}(Y_i^K))$	Kullback-Leibler divergence between current output distribution $P(Y_i^K)$ and desired target output distribution $\tilde{P}(Y_i^K)$ (constant target firing rate of 30Hz)
β, γ	optimization constants

We have derived learning rules which perform gradient ascent on the objective functions L^{IB} (3) and L_i^{IC} (4). The resulting update rules are an extension to the generalized Bienenstock-Cooper-Munro (BCM) rule for spiking neurons [3].

3.1 Spike-based learning rules

Performing gradient ascent on L^{IB} (3) and L_i^{IC} (4) yields online learning rules for the weights of neuron i, w_{ij}. The weight change Δw_{ij}^k at time $t^k = k\Delta t$ is given by

$$\frac{\Delta w_{1j}^k}{\Delta t} = -\alpha C_{1j}^k \left[B_1^k(-\gamma) - \beta \Delta t B_{12}^k \right] \qquad \text{for the IB case, and} \tag{5}$$

$$\frac{\Delta w_{ij}^k}{\Delta t} = \alpha C_{ij}^k \left[B_i^k(\gamma) - \beta \Delta t B_{12}^k \right] \qquad \text{for the IC case,} \tag{6}$$

with a learning rate $\alpha > 0$ and optimization parameters β and γ.

The **correlation term** C_{ij}^k measures coincidences between postsynaptic spikes at neuron i and PSPs generated by presynaptic action potentials arriving at synapse j:

$$C_{ij}^k = C_{ij}^{k-1} \left(1 - \frac{\Delta t}{\tau_C}\right) + \sum_{n=1}^{k} \epsilon(t^k - t^n) x_j^n \frac{g'(u_i(t^k))}{g(u_i(t^k))} \left[y_i^k - \rho_i^k \right] \tag{7}$$

τ_C	time constant of exponential correlation window (1s)
y_i^k	binary variable indicating an output spike of neuron i in the k-th time step
$g'(u)$	derivative of $g(u)$ with respect to u

The term B_i^k is responsible for optimizing the **mutual information between input and output** and maintaining the constant target firing rate for neuron i. It compares the current firing rate $g(u_i(t^k))$ with its running average $\bar{g}_i(t^k)$, and simultaneously the running average $\bar{g}_i(t^k)$ with the constant target rate $\tilde{g}$:

$$\begin{aligned} B_i^k(\gamma) = {} & \frac{y_i^k}{\Delta t} \log \left[\frac{g(u_i(t^k))}{\bar{g}_i(t^k)} \left(\frac{\tilde{g}}{\bar{g}_i(t^k)} \right)^{\gamma} \right] \\ & - (1 - y_i^k) R_i(t^k) \left[g(u_i(t^k)) - (1 + \gamma)\bar{g}_i(t^k) + \gamma \tilde{g} \right] \end{aligned} \tag{8}$$

The term B_{12}^k measures the **mutual information between spike trains** Y_1^K and Y_2^K. It basically compares the average product of firing rates $\bar{g}_{12}(t^k)$ with the product of average firing rates $\bar{g}_1(t^k)\bar{g}_2(t^k)$:

$$\begin{aligned} B_{12}^k = {} & \frac{y_1^k y_2^k}{(\Delta t)^2} \log \frac{\bar{g}_{12}(t^k)}{\bar{g}_1(t^k)\bar{g}_2(t^k)} - \frac{y_1^k}{\Delta t}(1 - y_2^k) R_2(t^k) \left[\frac{\bar{g}_{12}(t^k)}{\bar{g}_1(t^k)} - \bar{g}_2(t^k) \right] \\ & - \frac{y_2^k}{\Delta t}(1 - y_1^k) R_1(t^k) \left[\frac{\bar{g}_{12}(t^k)}{\bar{g}_2(t^k)} - \bar{g}_1(t^k) \right] \\ & + (1 - y_1^k)(1 - y_2^k) R_1(t^k) R_2(t^k) \left[\bar{g}_{12}(t^k) - \bar{g}_1(t^k)\bar{g}_2(t^k) \right] \end{aligned} \tag{9}$$

3.2 Rate-based learning rule for Information Bottleneck

For a simplified neuron model without refractoriness the spike-based rule for the IB case (5) reduces to the following rate-based rule:

$$\begin{aligned} \frac{\Delta w_{1j}^k}{\Delta t} = -\alpha \nu_j^{pre,k} f(\nu_1^k) \Bigg\{ & \log \left[\frac{\nu_1^k}{\bar{\nu}_1^k} \left(\frac{\bar{\nu}_1^k}{\tilde{g}} \right)^{\gamma} \right] \\ & - \beta \Delta t \left(\nu_2^k \log \left[\frac{\bar{\nu}_{12}^k}{\bar{\nu}_1^k \bar{\nu}_2^k} \right] - \bar{\nu}_2^k \left[\frac{\bar{\nu}_{12}^k}{\bar{\nu}_1^k \bar{\nu}_2^k} - 1 \right] \right) \Bigg\} \end{aligned} \tag{10}$$

$\nu_j^{pre,k}$	presynaptic firing rate at synapse j at time t^k
$f(\nu_1^k)$	sensitivity of neuron 1 at its current firing state ν_1^k
ν_1^k	output firing rate of neuron 1 at time t^k
ν_2^k	firing rate of the target signal at time t^k
$\bar{\nu}_1^k$, $\bar{\nu}_2^k$	running averages of ν_1^k and ν_2^k
$\bar{\nu}_{12}^k$	running average of the product $\nu_1^k \nu_2^k$

4 Relation to the BCM rule

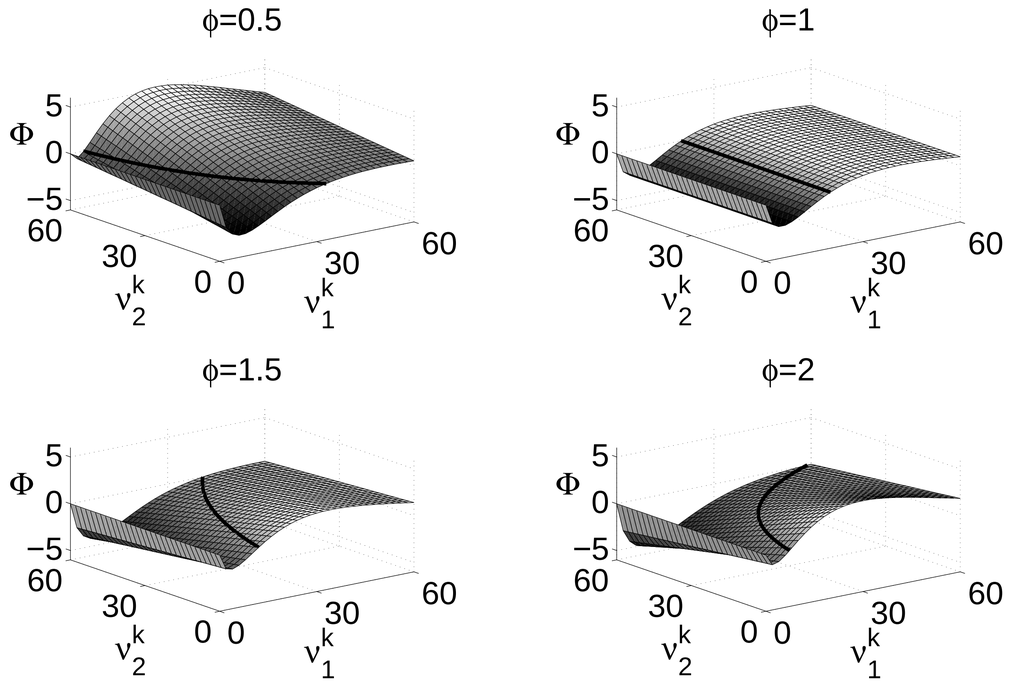
We can rewrite the rate-based learning rule (10) as

$$\frac{\Delta w_{1j}^k}{\Delta t} = -\alpha \nu_j^{pre,k} \Phi(\nu_1^k, \nu_2^k) \tag{11}$$

and view it as an extension of the classical Bienenstock-Cooper-Munro (BCM) rule [5] with a two-dimensional synaptic modification function $\Phi(\nu_1^k, \nu_2^k)$. Values of $\Phi > 1$ produce LTD whereas values of $\Phi < 1$ produce LTP. These regimes are separated by a sliding threshold that does not only depend on the running average of the postsynaptic rate $\bar{\nu}_1^k$, but also on the current values of ν_2^k and $\bar{\nu}_2^k$.

Institute for Theoretical Computer Science,
Graz University of Technology,
A-8010 Graz, Austria

Extension of the classical BCM rule to 2 dimensions:



This figure shows the function $\Phi(\nu_1^k, \nu_2^k)$ for different values of $\phi = \bar{\nu}_{12}^k/(\bar{\nu}_1^k\bar{\nu}_2^k)$ and for the special case $\bar{\nu}_1^k = \bar{\nu}_2^k = \tilde{g} = 20$Hz. For $\phi = 1$ (top right) it reduces to a one-dimensional function, as in the classical BCM-rule. In each plot the black solid line indicates the transition from depression to potentiation ($\Phi = 0$).

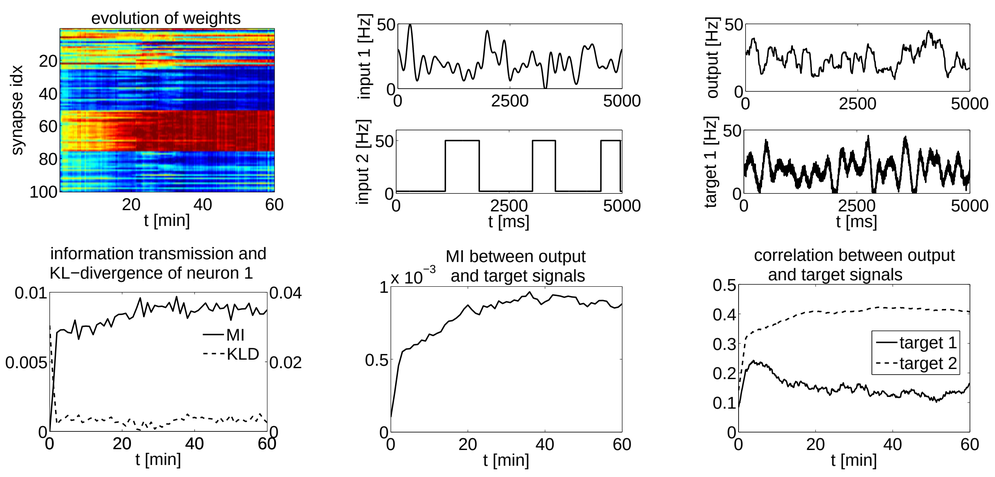
5 Results (Information Bottleneck)

We maximize the information between the output Y_1^K of a learning neuron and two target signals, Y_2^K and Y_3^K, and get the learning rule

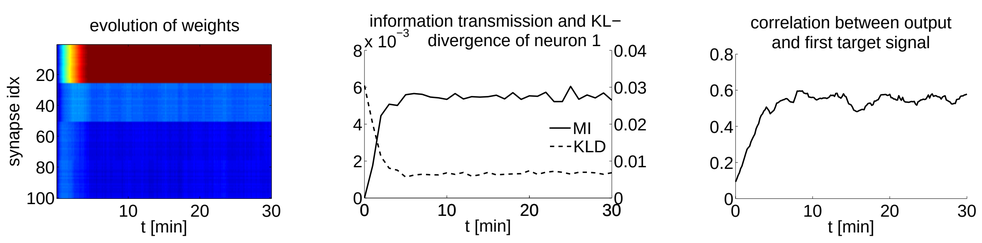
$$\frac{\Delta w_{1j}^k}{\Delta t} = -\alpha C_{1j}^k \left[B_1^k(-\gamma) - \beta \left(B_{12}^k + B_{13}^k \right) \right], \tag{12}$$

where B_{12}^k and B_{13}^k are terms characterizing the statistical dependence between the output Y_1^K and target signals Y_2^K and Y_3^K, respectively (cf. equ. (9)).

- We use two different kinds of target signals: one which has a similar rate modulation to one part of the input, and one which has spike-spike correlations with another part of the input.
- 100 synapses are divided into 4 groups: The first two groups receive rate modulated Poisson spike trains, the other two groups receive correlated spike trains at a constant rate (20Hz, $cc = 0.5$). Spike trains from different groups are uncorrelated.
- The first target signal, Y_2^K, has a similar rate modulation as input group 1, and the second target spike train, Y_3^K, is correlated with inputs from group 3. Both target signals are silent during random intervals.



Strong weights are developed for those parts of the input which are correlated to one of the target signals (groups 1 and 3). The output rate approximates the rate of the target signal. The mutual information between output and target signal increases, whereas the information between input and output is kept as low as possible.



With the rate-based learning rule (10) strong weights grow only for input group 1. It is not able to detect spike-spike correlations between outputs and the target signals.

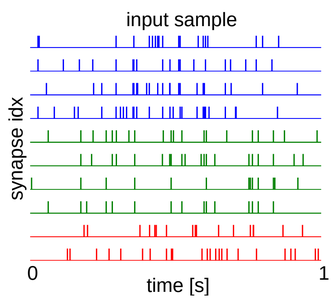
6 Results (Independent Components)

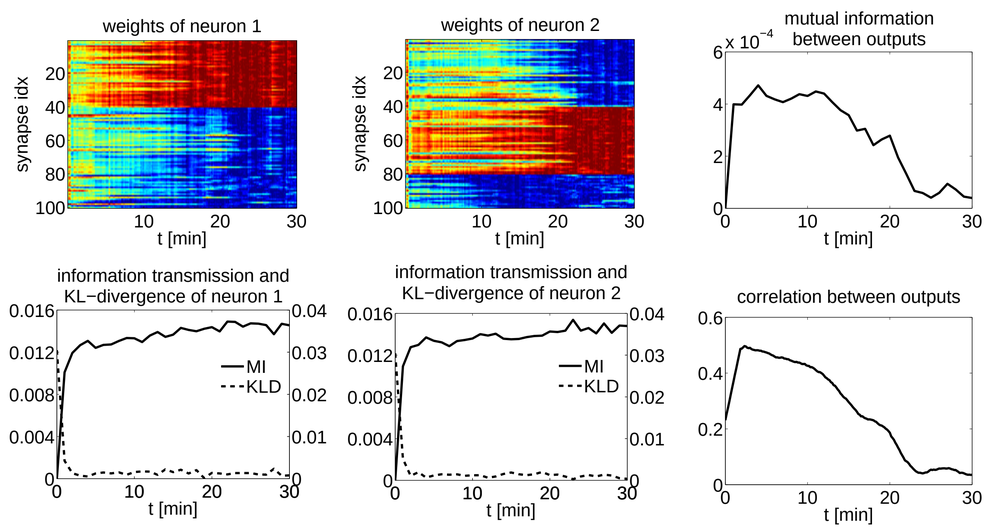
We use a biologically more realistic setup where the the nonlocal term B_{12}^k (9) is implemented by interneurons IN1 to IN4 that modulate the gain function $g(u_i(t))$. The gain $g(u_1(t^k))$ of neuron 1 is modified by IN2 and IN4 according to

$$\hat{g}_1(t^k) = \begin{cases} g(u_1(t^k)) \exp\left[R_2(t^k)\beta\Delta t \left(\frac{\bar{g}_{12}(t^k)}{\bar{g}_1(t^k)} - \bar{g}_2(t^k) \right) \right] & \text{if neuron 1 has spiked (IN4),} \\ g(u_1(t^k)) - \tilde{\beta} \left[\frac{\bar{g}_{12}(t^k)}{\bar{g}_2(t^k)} - \bar{g}_1(t^k) \right] & \text{if neuron 2 has spiked (IN2),} \end{cases} \tag{13}$$

where the parameter $\tilde{\beta}$ is scaled such that $g_1(t^k)$ does not become negative. In the absence of spikes $\hat{g}_1(t^k)$ decays back exponentially to the original $g(u_1(t^k))$. The gain $g(u_2(t^k))$ of neuron 2 is changed in a symmetric way (by interneurons IN1 and IN3).

- Two neurons receive the same input at 100 synapses, consisting of constant rate Poisson spike trains (20Hz).
- The input is divided into two groups of 40 correlated spike trains each ($cc = 0.5$); spike trains from different input groups are uncorrelated.
- The remaining 20 synapses receive uncorrelated Poisson input.





Each neuron develops strong weights for a different correlated group. The information transmission of both neurons is maximized, whereas the mutual information between the outputs decreases.

7 Discussion

Information Bottleneck (IB) and Independent Component Analysis (ICA) have been proposed as principles for unsupervised learning in lower cortical areas, however, learning rules that can implement these principles with spiking neurons have still been missing. In this work we have derived learning rules for such tasks for a stochastically spiking neuron with refractoriness from information theoretic principles. Furthermore, we have demonstrated that the extraction of independent components can be implemented in a biologically realistic manner, using inhibitory interneurons for gain control.

References

[1] N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. In *Proceedings of the 37-th Annual Allerton Conference on Communication, Control and Computing*, pages 368–377, 1999.

[2] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley, New York, 2001.

[3] T. Toyoizumi, J.-P. Pfister, K. Aihara, and W. Gerstner. Generalized Bienenstock-Cooper-Munro rule for spiking neurons that maximizes information transmission. *Proc. Natl. Acad. Sci. USA*, 102:5239–5244, 2005.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

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

[6] H. Markram, M. Toledo-Rodriguez, Y. Wang, A. Gupta, G. Silberberg, and C. Wu. Interneurons of the neocortical inhibitory system. *Nat Rev Neurosci.*, 5(10):793–807, 2004.

[7] W. Gerstner and W. M. Kistler. *Spiking Neuron Models*. Cambridge University Press, Cambridge, 2002.