



HAL
open science

Sparse Hebbian Learning is efficient with egalitarian homeostasis

Laurent Perrinet

► **To cite this version:**

Laurent Perrinet. Sparse Hebbian Learning is efficient with egalitarian homeostasis. 2007. hal-00156610v2

HAL Id: hal-00156610

<https://hal.science/hal-00156610v2>

Preprint submitted on 5 Sep 2007 (v2), last revised 7 Dec 2016 (v7)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sparse Hebbian Learning is efficient with egalitarian homeostasis

Laurent U. Perrinet*

Institut de Neurosciences Cognitives de la Méditerranée (INCM)

CNRS / University of Provence

13402 Marseille Cedex 20, France

e-mail: Laurent.Perrinet@incm.cnrs-mrs.fr

September 5, 2007

Abstract

Following the work of Olshausen and Field [1998], a number of unsupervised learning algorithms were proposed for receptive field formation in the primary visual cortex (V1). We describe here the theoretical formulation of a novel and simple algorithm aiming at achieving the shortest representation in an over-complete dictionary of unknown features using the L_0 norm, a “hard” NP-complete problem. Inspired by parallel computing solutions with spiking events, we propose here a solution based on correlation-based inhibition with a tuned egalitarian homeostasis. We present results of the simulation of this dynamical neural network with natural images and compare it to the SPARSENET solution. This algorithm exhibited similarly the formation of edge-like components as is observed in the input layer of V1 and to assess the quality of the different sets of filters, we additionally compared their relative efficiencies and show that our sub-optimal solution to the hard problem performs better than the optimal solution to the relaxed problem. We also show that homeostasis, by tuning the competition and cooperation, may yield solutions of different qualities which coding efficiency drastically depended on the goal assigned to the model by the cost function. This model provides a bridge between the non-linearity of the neural response and optimal use of distributed probabilistic representations of information. It is suggesting the importance of the role of interactions within neural assemblies [Hebb, 1949] to efficiently build representations in the neural code.

Keywords

Neural code, spike-event computation, correlation-based inhibition, Adaptive Matching Pursuit, Sparse-Hebbian Learning

*See supplementary material on <http://incm.cnrs-mrs.fr/LaurentPerrinet/SparseHebbianLearning>.

1 Sparse Hebbian Learning : Efficient coding and learning in V1

It is still largely unclear how information is processed in the central nervous system and in particular how the sensory visual system allows for an efficient and robust perception, in particular how gain control mechanisms are implemented so that information from different modalities, possibly ambiguous, may be efficiently integrated. Moreover, a main challenge in neuroscience is to understand how the central nervous system adapts to the statistics of this information and learns from a fairly unorganized network at birth to a structure which, as we grow up, is efficiently adapted to the ever changing sensory world. The recent successes of explaining neural responses by Bayesian processes [Albrecht and Geisler, 1995; Mamassian, 2002] often supposes that the neural activity represents a solution of an inverse problem that would “reverse engineer” the world and that neurons achieve some sort of statistical inference through a stochastic approximation. However, to understand how neural computations operate, we lack a theory of how to relate this function to the structure and dynamics of neural networks. This fact is striking in the Spike-Time Dependent Plasticity (STDP) [Abbott and Nelson, 2000] and though some functional approaches do exist [Bell and Parra, 2005; Toyozumi et al., 2007], a majority of algorithms blindly apply the learning rule which was isolated in experiments without investigating its function or efficiency. In [Guyonneau et al., 2005] for instance, the solution of the algorithm is proposed to “emerge” even though it is simply deterministic (see e. g. [Kempster et al., 1999] for a line of demonstration) and the extension of this algorithm in [Masquelier and Thorpe, 2007] is thus equivalent to an algorithm similar to Vector Quantization. Instead, we will propose to precisely define the function at hand to better explain the neural code as the link between structure (spikes in a distributed network) and function (efficient coding) and explore the significant parameters at work in these mechanisms.

As outlined in a number of works [Field, 1994; Barlow, 2001], a goal of neural computations is to build efficient intermediate representations to allow efficient decision making. In low-level sensory areas, we may formulate this as an inverse problem and use a *Linear Generative Model* (LGM) as the forward model [Olshausen and Field, 1998]. As in this paper, we chose here to restrict ourselves to study the selection of optimal filters on imagelets (that is small patches from natural images) in the framework defined by the SPARSENET algorithm. In particular, these images are static, grayscale and filtered according to similar parameters to allow a one-to-one comparison of the different algorithms. It is then expected that the resulting filters will help us to understand the processes underlying receptive field formation in the input layer of V1 (layer 4) but also on the computations (especially interactions) that occur at this level.

In order to define the LGM we will use a “dictionary” of sources as the matrix $\mathbf{A} = \{\mathbf{A}_j\}_{1 \leq j \leq N}$ of size N , with $\mathbf{A}_j = \{A_{ij}\}_{1 \leq i \leq M}$ over the set of sampling positions i (that is the pixels in a simple image processing framework). This dictionary is possibly larger than the dimension of the input space (that is when $N > M$); the dictionary is then said to be over-complete. Knowing \mathbf{A} and the corresponding sources

$\mathbf{s} = \{s_j\}_{1 \leq j \leq N}$, the signal $\mathbf{x} = \{x_i\}_{1 \leq i \leq M}$ is defined as

$$\mathbf{x} = \sum_{1 \leq j \leq N} s_j \cdot \mathbf{A}_j + \mathbf{n} = \mathbf{A} \cdot \mathbf{s} + \mathbf{n} \quad (1)$$

where \mathbf{n} is a gaussian additive noise which is decorrelated thanks to the preprocessing and of variance σ^2 . This is well adapted to natural scenes because transparency laws are linear for luminances and the LGM describes well the synthesis in a local neighborhood of any natural image. The goal of any coding algorithm is to find for an observed \mathbf{x} the best set of sources that generated the signal and for a learning algorithm to adapt at best in the long term to the parameters of the LGM, that is to the matrix \mathbf{A} . We may therefore theoretically assign the functionality for simple cells in layer 4 of V1 as adapting at best with an initialisation with random receptive fields to provide the most efficient coding of the inverse LGM model.

However, finding the “best” solution is subject to the the definition of the efficiency, that is to the choice of an objective function. Even then, many different solutions may exist for the same problem. For instance, the SPARSENET algorithm—which is a state-of-the-art algorithm for explaining the formation of edge-like receptive fields in V1—optimizes a measure of the *sparseness* of the representation by a LGM while keeping the representation accurate by defining a cost:

$$\mathcal{C} = \frac{1}{2\sigma^2} \cdot \|\mathbf{x} - \sum s_j \cdot \mathbf{A}_j\|^2 + \mathcal{S}(\mathbf{s}) \text{ with e.g. } \mathcal{S} = \lambda \cdot \|\mathbf{s}\|_1 = \lambda \cdot \sum |s_j| \quad (2)$$

where σ is the mean energy of images and λ a positive constant acting as a regularization parameter. This liberty in the definition of the sparseness \mathcal{S} leads to a wide variety of proposed solutions to sparse coding [Pece, 2002] such as optimization [Olshausen and Field, 1998; Lee et al., 2007], non negative matrix factorization [Lee and Seung, 1999; Ranzato et al., 2007] or by using Matching Pursuit [Smith and Lewicki, 2006; Rehn and Sommer, 2006]. Another set of algorithms allow the inclusion of transform groups to see the formation of invariant representations and topographic maps [Hyvärinen et al., 2001; Bednar et al., 2004]. These solutions to the unsupervised learning problem have similarities since they use this sparse efficient representation to perform then an Hebbian-type of learning. They are important both because they have numerous applications (such as providing novel algorithms for ICA) but also because they may help us to understand the function of neural computations for the unsupervised emergence of structures in V1.

The paper is organized as follows. We will first provide the theoretical framework of our algorithm based on the spiking events. This will highlight differences and similarities with the SPARSENET algorithm, in particular concerning the coding of information and the homeostatic rule. We will then show quantitative results for these algorithms and particularly how they compare in terms of coding cost and their stability to parameter changes. We will finally apply this method to show the importance of homeostasis in the quality of the resulting filters.

2 Method : Adaptive Matching Pursuit with egalitarian homeostasis

As in Occam’s razor, one may state that given two solutions of similar quality, the best is the one with lowest complexity. This can be formalized in a probabilistic framework using Shannon’s information of the solution \mathbf{s} given the model’s parameters ($\mathcal{C} = E(-\log P(\mathbf{s}|\mathbf{x}, \mathbf{A}))$), where $E(\cdot)$ denotes averaging over multiple images) by using Bayes’ theorem as the sum of its likelihood probability (or equivalently its entropic coding) of the set of sources added to the description length:

$$\log P(\mathbf{s}|\mathbf{x}, \mathbf{A}) = -\log Z + \left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \sum s_j \cdot \mathbf{A}_j\|^2\right) + \log P(\mathbf{s}|\mathbf{A}) \quad (3)$$

where Z is the partition function. This is equivalent to the sparseness cost defined in Eq. 2 if the *a priori* probability in the LGM model corresponds to a factorial Laplacian prior. Similarly, we will model the LGM by assuming independence of the coefficients but also use an “egalitarian homeostasis” hypothesis: in the context of neural coding, neurons are *a priori* equal and the prior of coefficients are all identical. Moreover, spikes are all-or-none events and if they carry a binary representation [Dewese and Zador, 2003] from a dictionary of size M , the cost may then be defined as:

$$\mathcal{C} = \frac{1}{2\sigma^2} \cdot \|\mathbf{x} - \sum_j s_j \cdot \mathbf{A}_j\|^2 + \log_2(M) \cdot \|\mathbf{s}\|_0 + \mathcal{Q}(\mathbf{s}) \quad (4)$$

where $\|\mathbf{s}\|_0$ is the length of the retrieved solution (or also the L_0 norm) and $\mathcal{Q}(\mathbf{s})$ is the quantization cost, that is the error introduced by using approached coefficients and which was typically small. Note first that for any coding, this cost function is dynamic since the number of spikes may increase in time but also that it links efficiency to sparseness, as with information criterions such as the AIC [Akaike, 1974]. It also explicitly rates the economy of consumed metabolic resources as is used in [Rehn and Sommer, 2006], but we retain this only as a consequence of the algorithm. However, resolving the coding problem with the L_0 norm is NP-complete with respect to the dimension of the dictionary [Mallat, 1998, p. 409].

The *Sparse Spike Coding* (SSC) algorithm is inspired by work from Perrinet et al. [2002]; Perrinet [2004, 2007] and provides an efficient approximation to the coding problem which we will use here. It is a greedy approach applied on the efficiency criterion defined in Eq. 4 based on two repetitive steps. First, given the signal \mathbf{x} , we are searching for the *single* source $s_{j^*} \cdot \mathbf{A}_{j^*}$ that corresponds to the maximum *a posteriori* (MAP) realization for \mathbf{x} . It is defined by:

$$j^* = \text{ArgMax}_j [f_j(\langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{A}_j}{\|\mathbf{A}_j\|} \rangle)] \quad (5)$$

with $\langle \cdot, \cdot \rangle$ denoting the scalar product and $f_j(\cdot)$ is some gain function that we will describe below and which may be set initially to identical, strictly increasing functions. In a second step, the information is fed-back to correlated sources through :

$$\mathbf{x} \leftarrow \mathbf{x} - s_{j^*} \cdot \mathbf{A}_{j^*} \quad (6)$$

where $s_{j^*}^*$ is the scalar projection $\langle \mathbf{x}, \mathbf{A}_{j^*} / \|\mathbf{A}_{j^*}\|^2 \rangle$. Equivalently we may propagate laterally:

$$\langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{A}_j}{\|\mathbf{A}_j\|} \rangle \leftarrow \langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{A}_j}{\|\mathbf{A}_j\|} \rangle - \langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{A}_{j^*}}{\|\mathbf{A}_{j^*}\|} \rangle + \langle \frac{\mathbf{A}_{j^*}}{\|\mathbf{A}_{j^*}\|}, \frac{\mathbf{A}_j}{\|\mathbf{A}_j\|} \rangle \quad (7)$$

where $z_j = \frac{s_{j^*}^* \cdot \|\mathbf{A}_{j^*}\|}{\|\mathbf{x}\|} = \langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{A}_{j^*}}{\|\mathbf{A}_{j^*}\|} \rangle$ is the quality of the match measured by the normalized correlation, that is the multidimensional cosine (of norm inferior to 1). This correspond to a correlation-based inhibition as may be observed to be necessary for the formation of elongated receptive fields [Bolz and Gilbert, 1989]. The algorithm is then iterated with Eq. 5 until some stopping criteria is reached.

On a longer time scale, the adaptation of the system is as in SPARSENET is based similarly in slowly adaptating the dictionary to the signal thanks to the sparse solution given by the coding algorithm. We may implement this for every image at every coding step since we have an evaluation of the log-likelihood by the distance of the residual image to the selected filter, that is to $\|\mathbf{x} - s_{j^*}^* \cdot \mathbf{A}_{j^*}\|^2$, the rest being regarded as a perturbation which should cancel out by integrating it in time. At every step after Eq. 5 and using the gradient descent approach as in [Olshausen and Field, 1998], we infer that we may slowly modify the winning weight vector corresponding to the winning filter \mathbf{A}_{j^*} by taking it closer to $\frac{\mathbf{x}}{s_{j^*}^*}$:

$$\frac{\partial \mathcal{C}}{\partial \mathbf{A}_{j^*}} = \frac{\partial}{\partial \mathbf{A}_{j^*}} \frac{1}{2\sigma^2} \cdot \|\mathbf{x} - s_{j^*}^* \cdot \mathbf{A}_{j^*}\|^2 \quad (8)$$

$$= \frac{1}{2\sigma^2} \cdot s_{j^*}^* (\mathbf{x} - \sum s_{j^*}^* \cdot \mathbf{A}_{j^*}) \quad (9)$$

that is

$$\mathbf{A}_{j^*} \leftarrow \mathbf{A}_{j^*} + \eta \frac{\mathbf{x}}{s_{j^*}^*} \quad (10)$$

where η is the learning rate, inversely proportional to the time scale of the features learned¹. Similar approaches have been taken based on a similar correlation-based inhibition that we regroup under the name of Sparse-Hebbian Learning [Smith and Lewicki, 2006; Rehn and Sommer, 2006].

As it is, the algorithm (as well as SPARSENET) is unstable. In fact, since we start with random filters, it is more likely that a salient feature was selected first and will be selected with a higher probability in subsequent learning steps. An homeostatic regulation with a similar time-scale of the learning is therefore necessary to ensure convergence of the learning algorithm. Whereas SPARSENET uses the norm of the filters to control coefficients' distributions across neurones, the SSC matching criteria (see Eq. 5) is independent to the norm of the filters. We thus rather adapted the sensibility of neurones by using the $f_j(\cdot)$ functions and tuned them to the cumulative probability distribution by using the same mechanism decribed in [Laughlin, 1981; Atick, 1992] over the matching coefficients:

$$f_j(z_j^*) = \int^{z_j^*} dz_j P(z_j) \quad (11)$$

¹This hebbian rule [Hebb, 1949] is similar to Eq. 17 in [Olshausen and Field, 1998] or to Eq. 2 in [Smith and Lewicki, 2006].

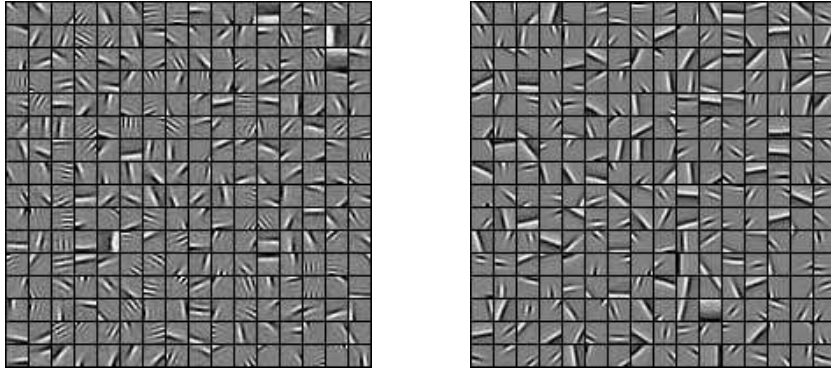


Figure 1: **Results of the proposed SHL scheme compared to SPARSENET**. Starting with random filters, we compare here the results of the learning schemes with 256 filters at convergence (2000 steps) using **(Left)** the classical conjugate gradient function method as is used in [Olshausen and Field, 1998] with **(Right)** the Sparse Spike Coding method. Filters of the same size as the imagelets (12×12) are presented in a matrix (separated with a black border). Note that their position in the matrix is as in ICA arbitrary. Results replicate the original results of [Olshausen and Field, 1998] and are similar for both methods: both dictionary consist of gabor-like filters which are similar to the receptive fields of simple cells in the primary visual cortex. Edges appear in these conditions to be the independent components of natural images.

This method ensures that the probability of choosing a neurone at any time was uniform and identical, hence that the cost function that we draw from the egalitarian homeostasis hypothesis was valid. This mechanism provided to be robust and qualitatively important for the convergence of the algorithm.

3 Results : comparison with SPARSENET and importance of homeostasis

We compared this novel *Sparse Hebbian Learning* algorithm with the SPARSENET algorithm. In fact, this algorithm as other similar schemes mainly differs by the coding method used to obtain the sparse representation. We particularly focused in this paper in the validation and quantitative comparison of both algorithms in terms of efficiency on the task we defined². We used a similar context and architecture as the experiments described in [Olshausen and Field, 1998] and used in particular the database of inputs of the SPARSENET algorithm. Here, we show the results for 12×12 patches (so that

²The whole collection of simulation scripts were written with the intention of controlling the convergence of the algorithms and the effect of the different parameters. All scripts to reproduce the figures and supplementary material are available on the author's website at <http://incm.cnrs-mrs.fr/LaurentPerrinet/SparseHebbianLearning>. Version 1.5 and experiment 20061018T200936 was used for this paper, and other figures regarding control experiments may be found there. The original parameters of SPARSENET were used for the CGF algorithm.

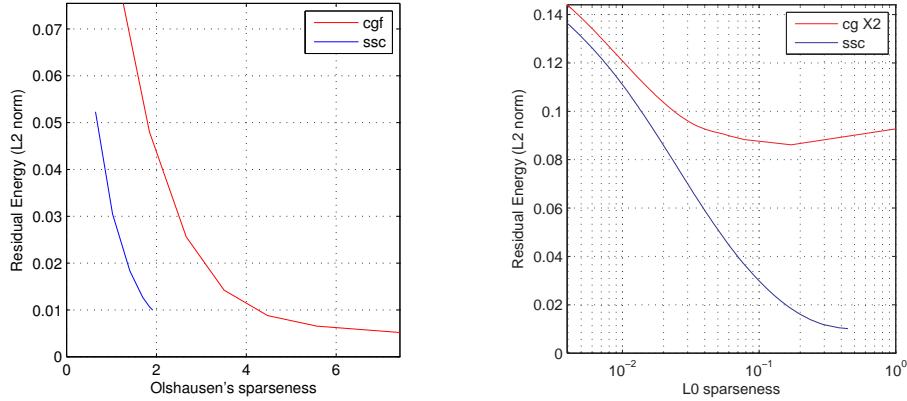


Figure 2: **Efficiency of the proposed SHL scheme compared to SPARSENET** . We evaluated the quality of the SHL algorithm with two different coding strategies by comparing the coding efficiency of the sparse spike coding ('ssc') method with the classical conjugate gradient function ('cgf') method as is used in [Olshausen and Field, 1998] for the coding of a set of 5000 image patches drawn from a database of natural images. After convergence of the learning phase (see Fig. 1), we plot the mean final residual error (L_2 norm) as a function of two definitions of sparseness: (*Left*) the sparseness of the coefficients (as defined in [Olshausen and Field, 1998]) and (*Right*) the relative number of active (or non-zero) coefficients (that is the normalized L_0 norm and the coding step for SSC) and which provides an estimate of the mean coding efficiency for the image patches. Best results are those giving a lower error for a given sparsity or a lower sparseness (better compression) for the same error. Occam's razor translates in this figure into the fact for a given L_2 norm, the L_0 norm is lower (an horizontal line would cross from left the best solution first). In both cases, the proposed algorithm provides a paradigm which is of better efficiency compared to SPARSENET . It should be noted that it is also superior for the cost based on the L_1 norm, a result which may reflect the fact that the L_0 norm defines a stronger sparseness constraint.

$M = 144$) from the whitened images and we chose to learn 256 filters. Results show the emergence of edge-like filters (see Fig. 1) for a wide range of parameters. It should be noted that the output of both coding algorithms give non-linear results for the mixing of images (the output to the sum of two images is not necessarily the sum of both individual output). Moreover, the response to rotated filters will yield a response (the selectivity curve) which is typically sharper than the linear response [Perrinet, 2005]. This provides a simple model for orientation selectivity and in particular will exhibit the same non-linearity thanks to SSC without adding a non-linear gain control to match physiological recordings [Carandini et al., 2005]. However, it is not clear by the shape of the filters alone which solution is most efficient.

To address this question, we used a combined approach to assess the quality of the resulting filters. Classical methods evaluate this by a qualitative analysis of the filters' shape, by fitting them with Gabor filters [Lewicki and Sejnowski, 2000] or by com-

paring the distribution of filters with neurophysiological experiments. To quantify the efficiency of both solutions, we rather compared the efficiency of both algorithms using the learned basis for both costs:

1. First, by changing an internal parameter which is tuning the compromise between reconstruction error and the sparsity (namely the estimated variance of the noise for the conjugate gradient method and the stopping criterion in the pursuit), one could yield different mean residual error with different mean sparseness of the coefficients, as defined in Eq. 2 (see Fig. 2, left).
2. In a second experiment, we compared the efficiency of the greedy pursuit while varying the number of active coefficients (the L_0 norm), that is the rank of the pursuit or the number of spikes (Eq. 4). To compare this method with the conjugate gradient, a first pass of the latter method was assigning for a fixed number of active coefficients the best neurones while a second pass optimized the coefficients for this set of "active" vectors (see Fig. 2, right).

Controlling with a wide range of parameters and a variety of methods yielded similar qualitative results (such as changing the learning rate, the parameters of the conjugate gradient or using the natural gradient) proving that the hebbian learning converged robustly as long as the coding algorithm provided a good sparse representation of the input. As a result, it appeared in a robust manner that the greedy solution to the hard problem (that is SHL) is more efficient for the optimized cost but also to the cost defined in the relaxed problem.

The choice of the homeostatic regulation was based on the cost function and the hypothesis that led to it. In fact, by assuming that all neurons should be chosen with equal probability, we impose a strong constraint for the neural assembly (all neurons should be equal). On the other hand, when choosing a more relaxed system (such as using that the filters to be normalized or by using the homeostatic rule defined in SPARSENET) we obtain qualitatively different filters whose quality would depend on the chosen cost function. We therefore performed an analysis for the SHL scheme with SCC for the different homeostasis schemes (see Fig. 3, Left). As expected, the homeostasis algorithm performed more efficiently (see Fig. 3, Right) and an heuristic approach (by approaching the f_j functions) yielded similar results which show the efficiency of such a method compared to the method without homeostasis as was used in other schemes such as [Smith and Lewicki, 2006]. In fact, in this relaxed algorithm the filters will correspond to features of different saliencies. In particular, the 'textural' filters due to their lower generality, will be more likely to be selected with lower z_j coefficients and correspond more to the Fourier filters (that one may obtain by PCA or the simple hebbian rule) that are still optimal to code arbitrary imagelets such as noise [Li, 2006]. The SHL algorithm ensures that all will be selected equally and thus that textured elements will be less probable. However, when increasing the over-completeness, one sees that these filters appear along with the over representation of salient features (a similar edge with different phases for instance). Exploring the results for different dimensions of the dictionary may give an evaluation of the optimal complexity of the LGM to describe imagelets in terms of a trade-off between accuracy and generality.

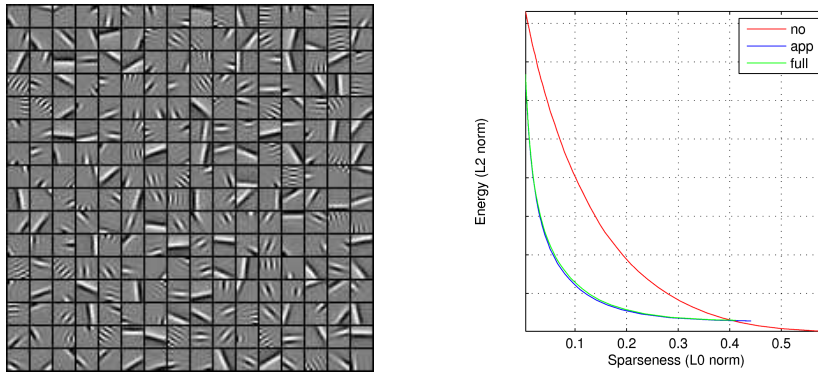


Figure 3: **Role of homeostasis.** (*Left*) When relaxing the homeostatic constraint in the SHL algorithm to the one implemented in SPARSENET , the algorithm converges to a set of filters which contains some less localized filters and to some high-frequency Gabors which correspond to more 'textural' features. One may wonder if these filters are inefficient and capturing noise or if they correspond to inherent features of natural images in this LGM model (see Fig. 2). (*Right*) The original set of filters gives a better result when using the L_0 norm for the proposed model based on a full probability model of the probability of coefficients ('full') but also for an approached model using an heuristic ('app'). One should note that the SHL method without egalitarian homeostasis ('no') gave better results after a certain size of the coefficient's vector since it also better represented textural and noise features.

Discussion

The work presented here is part of a larger program to assess qualitatively the efficiency of different solutions to computational neuroscience problems. We proved here that SSC proved superior to the SPARSENET architecture. Computationally, the complexity of the algorithms and the time required by both methods was similar on the different simulations on a computer (with a two fold advantage for SSC). However, the SSC algorithm consists of simpler operations (integrating and spiking) particularly adapted to an implementation on parallel architecture such as an aVLSI. It also provides a progressive result while the conjugate gradient method had to be recomputed for any different number of coefficients. We proved also that homeostasis played a significant role in these results and that counter-intuitively textured filters could also be good candidates for optimal coding in V1.

As a conclusion, we provided an original method for learning efficient representations using spiking neurones and based on neural computations. It offers a new interpretation for the receptive fields of neurones which in this view self organize in accordance with neighboring neurones to code at best the input and gives a caricatural yet simple model for orientation selectivity [Perrinet, 2005].

References

- Larry F. Abbott and Sacha B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–83, 2000. URL <http://www.nature.com/cgi-taf/DynaPage.taf?file=/neuro/journal/v3/n11s/full/n>
- H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–23, 1974.
- W. S. Albrecht and D. G. Geisler. Bayesian Analysis of Identification Performance in Monkey Visual Cortex: Nonlinear Mechanisms and Stimulus Certainty. *Vision Research*, 35(19):2723–30, 1995.
- Joseph J. Atick. Could Information Theory Provide an Ecological Theory of Sensory Processing? *Network: Computation in Neural Systems*, 3(2):213–52, 1992. URL <http://ib.cnea.gov.ar/~redneu/atick92.pdf>.
- Horace B. Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–25, 2001.
- James A Bednar, Amol Kelkar, and Risto Miikkulainen. Scaling self-organizing maps to model large cortical networks. *Neuroinformatics*, 2(3):275–302, 2004.
- Anthony J. Bell and Lucas C. Parra. Maximising information yields spike timing dependent plasticity. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 121–8. MIT Press, 2005.

- J Bolz and CD Gilbert. The role of horizontal connections in generating long receptive fields in the cat visual cortex. *European Journal of Neuroscience*, 1(3):263–8, 1989.
- Matteo Carandini, Jonathan B. Demb, Valerio Mante, David J. Tolhurst, Yang Dan, Bruno A. Olshausen, Jack L. Gallant, and Nicole C. Rust. Do we know what the early visual system does? *Journal of Neuroscience*, 25(46):10577–97, Nov 2005. doi: 10.1523/JNEUROSCI.3726-05.2005. URL <http://dx.doi.org/10.1523/JNEUROSCI.3726-05.2005>.
- M. Deweese and A. Zador. Binary coding in auditory cortex. *Journal of Neuroscience*, 23(21), August 2003.
- David J. Field. What is the goal of sensory coding? *Neural Computation*, 6(4):559–601, 1994.
- Rudy Guyonneau, Rufin VanRullen, and Simon J. Thorpe. Neurons tune to the earliest spikes through stdp. *Neural Computation*, 17(4):859–79, 2005. URL <http://neco.mitpress.org/cgi/content/abstract/17/4/859>.
- Donald O. Hebb. *The organization of behavior: A neuropsychological theory*. Wiley, New York, 1949.
- A. Hyvärinen, P. O. Hoyer, and M. Inki. Topographic Independent Component Analysis. *Neural Computation*, 13(7):1527–58, 2001.
- Richard Kempter, Wulfram Gerstner, and J. Leo van Hemmen. Hebbian Learning and Spiking Neurons. *Physical Review Letters*, E 59:4498–514, 1999.
- S. B. Laughlin. A simple coding procedure enhances a neuron’s information capacity. *Z. Naturforsch.*, 9–10(36):910–2, 1981.
- D. D. Lee and H. S. Seung. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature*, 401:788–91, 1999.
- Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Ng. Efficient sparse coding algorithms. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*, pages 801–808. MIT Press, Cambridge, MA, 2007.
- Michael S. Lewicki and Terrence J. Sejnowski. Learning Overcomplete Representations. *Neural Computation*, 12(2):337–65, 2000. URL citeseer.nj.nec.com/lewicki98learning.html.
- Zhaoping Li. Theoretical understanding of the early visual processes by data compression and data selection. *Network*, 17(4):301–334, 2006.
- Stéphane Mallat. *A wavelet tour of signal processing*. Academic Press, 1998.
- Pascal Mamassian. *Probabilistic Models of the Brain: Perception and Neural Function*, chapter Bayesian modelling of visual perception, pages 13–36. MIT Press, 2002.

- T Masquelier and SJ Thorpe. Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, 3(2):e31, 2007. doi: 10.1371/journal.pcbi.0030031.
- Bruno A. Olshausen and David J. Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? *Vision Research*, 37:3311–25, 1998.
- Arthur E. C. Pece. The problem of sparse image coding. *Journal of Mathematical Imaging and Vision*, 17:89–108, 2002.
- Laurent U. Perrinet. Feature detection using spikes : the greedy approach. *Journal of Physiology (Paris)*, 98(4-6):530–9, July-November 2004. doi: 10.1016/j.jphysparis.2005.09.012. URL <http://hal.archives-ouvertes.fr/hal-00110801/en/>.
- Laurent U. Perrinet. Efficient Source Detection Using Integrate-and-Fire Neurons. In W. Duch et al., editor, *ICANN 2005, LNCS 3696*, Lecture Notes in Computer Science, pages 167–72, Berlin Heidelberg, 2005. Springer. URL <http://incm.cnrs-mrs.fr/perrinet/publi/perrinet05icann.pdf>.
- Laurent U. Perrinet. *Topics in Dynamical Neural Networks: From Large Scale Neural Networks to Motor Control and Vision*, volume 142 of *The European Physical Journal (Special Topics)*, chapter Dynamical Neural Networks: modeling low-level vision at short latencies, pages 163–225. Springer Berlin / Heidelberg, mar 2007. doi: 10.1140/epjst/e2007-00061-7. URL <http://dx.doi.org/10.1140/epjst/e2007-00061-7>.
- Laurent U. Perrinet, Manuel Samuelides, and Simon Thorpe. Sparse spike coding in an asynchronous feed-forward multi-layer neural network using Matching Pursuit. *Neurocomputing*, 57C:125–34, 2002. URL <http://incm.cnrs-mrs.fr/LaurentPerrinet/Publications/Perrinet02sparse>. Special issue: New Aspects in Neurocomputing: 10th European Symposium on Artificial Neural Networks 2002 - Edited by T. Villmann.
- M. Ranzato, C.S. Poultney, S. Chopra, and Y. LeCun. Efficient learning of sparse overcomplete representations with an energy-based model. In Scholkopf et al., editor, *Advances in neural information processing systems*, volume 19. The MIT Press, Cambridge, MA, 2007.
- M Rehn and FT Sommer. A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *Journal of Computational Neuroscience*, 2006. doi: 10.1007/s10827-006-0003-9.
- E. Smith and M. S Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–82, 2006.
- Taro Toyozumi, Jean-Pascal Pfister, Kazuyuki Aihara, and Wulfram Gerstner. Optimality model of unsupervised spike-timing-dependent plasticity: synaptic memory and weight distribution. *Neural Computation*, 19(3):639–71, 2007. doi: 10.1162/neco.2007.19.3.639.