OPEN ACCESS Freely available online
PLOS COMPUTATIONAL BIOLOGY

Optimal Properties of Analog Perceptrons with Excitatory Weights

Claudia Clopath[1,2]*, **Nicolas Brunel**[1,3]

1 Laboratory of Neurophysics and Physiology, CNRS and Université Paris Descartes, Paris, France, **2** Centre for Theoretical Neuroscience, Columbia University, New York, New York, United States of America, **3** Departments of Statistics and Neurobiology, University of Chicago, Chicago, Illinois, United States of America

Abstract

The cerebellum is a brain structure which has been traditionally devoted to supervised learning. According to this theory, plasticity at the Parallel Fiber (PF) to Purkinje Cell (PC) synapses is guided by the Climbing fibers (CF), which encode an 'error signal'. Purkinje cells have thus been modeled as perceptrons, learning input/output binary associations. At maximal capacity, a perceptron with excitatory weights expresses a large fraction of zero-weight synapses, in agreement with experimental findings. However, numerous experiments indicate that the firing rate of Purkinje cells varies in an analog, not binary, manner. In this paper, we study the perceptron with analog inputs and outputs. We show that the optimal input has a sparse binary distribution, in good agreement with the burst firing of the Granule cells. In addition, we show that the weight distribution consists of a large fraction of silent synapses, as in previously studied binary perceptron models, and as seen experimentally.

Citation: Clopath C, Brunel N (2013) Optimal Properties of Analog Perceptrons with Excitatory Weights. PLoS Comput Biol 9(2): e1002919. doi:10.1371/journal.pcbi.1002919

Editor: Olaf Sporns, Indiana University, United States of America

Received September 28, 2012; **Accepted** December 27, 2012; **Published** February 21, 2013

Funding: This work has been supported by the Agence Nationale de la Recherche, grant ANR-08-SYSC-005 and by the Swiss National Science Foundation, grant PA00P3_139703. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cc3450@columbia.edu

Introduction

Purkinje cells (PCs) are the only outputs of the cerebellar cortex, a brain structure involved in motor learning. They receive a very large number ($\sim$150,000) of excitatory synaptic inputs from Granule Cells (GCs) through parallel fibers (PFs), and a single very strong input from the inferior olive through climbing fibers (CFs).

Single PCs have long been considered as a neurobiological implementation of a perceptron [1,2], the simplest feedforward network endowed with supervised learning [3], since CFs are thought to provide PCs with an error signal [4]. A perceptron learns associations between input patterns and a binary output that are imposed to it. Learning is due to synaptic modifications, under the control of an error signal. The learning capabilities of perceptrons have been extensively studied for unbiased [5,6] as well as biased patterns [6], and for unconstrained synapses [5,6]. In real neurons, synapses are either excitatory (glutamatergic synapses), or inhibitory (GABAergic synapses), depending on the identity of the pre-synaptic neurons (except during early development, when GABAergic synapses are initially excitatory and then become inhibitory). A multitude of experiments characterizing synaptic plasticity have shown that the strength, but not the sign, of a synapse can be modified by patterns of neuronal activity. This has led to the study of perceptrons with sign-constrained weights [7,8,9,10]. In particular, Brunel et al. [10] showed that when synaptic weights are constrained to be excitatory (positive or zero), a perceptron at maximal capacity has a distribution of synaptic weights with two components: a finite fraction of zero-weight ('silent') synapses; and a truncated Gaussian distribution for the rest of the synapses. They further showed that this distribution is in striking agreement with experimental data [10].

Numerous experiments show however that in the course of specific motor tasks, the firing rate of Purkinje cell varies in an analog, not binary, fashion [11,12,13,14]. We therefore set out to investigate the capacity and distribution of synaptic weights of a perceptron storing associations between analog inputs and outputs. More precisely, each input or output unit can take an analog value drawn from a distribution with a given mean and variance. We show that the optimal input distribution matches the firing pattern of the Granule cells, and weight distribution at maximal capacity reproduces the experimental Parallel Fiber to Purkinje cell synaptic weight distribution.

Results

The analog perceptron

The perceptron consists of N inputs and one output. Both inputs and outputs take continuous values. We require this perceptron to learn a set of p prescribed random input-output associations, where the inputs G_i^μ $(i=1,\ldots,N,\ \mu=1,\ldots,p)$ are drawn randomly and independently from a distribution $\rho_{in}(G)$, with mean μ_G and standard deviation σ_G while the target outputs P_t^μ are drawn randomly and independently from a distribution $\rho_{out}(P)$ with mean μ_P and standard deviation σ_P. Note that since G_i^μ and P_t^μ represent firing rates of input and output cells, respectively, they must be non-negative quantities. In particular, $\mu_G>0$, $\mu_P>0$ represent the mean firing rates of granule/Purkinje cell, respectively. The output of the perceptron when a pattern μ is presented in input is given by

$$P^\mu=\phi\left[\frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N} w_i G_i^\mu-\theta N\right)\right], \qquad (1)$$

Author Summary

Learning properties of neuronal networks have been extensively studied using methods from statistical physics. However, most of these studies ignore a fundamental constraint in networks of real neurons: synapses are either excitatory or inhibitory, and cannot change sign during learning. Here, we characterize the optimal storage properties of an analog perceptron with excitatory synapses, as a simplified model for cerebellar Purkinje cells. The information storage capacity is shown to be optimized when inputs have a sparse binary distribution, while the weight distribution at maximal capacity consists of a large amount of zero-weight synapses. Both features are in agreement with electrophysiological data.

where ϕ is a monotonically increasing static transfer function (f-I curve), w_i are the synaptic weights from input $i=1,\ldots,N$, θN represents inhibitory inputs that cancel the leading order term in $\sum_{i=1}^{N} w_i G_i^\mu$ so that the argument of ϕ is of order 1. In Purkinje cells, these inhibitory inputs are provided by interneurons of the molecular layer. The goal of perceptron learning is to find a set of synaptic weights $\{w_i \geq 0\}, i=1,\ldots,N$ for which $P^\mu = P_t^\mu$ for all $\mu=1,\ldots,p$.

We focus for simplicity on a linear transfer function $\phi(x)=x$, but our results can be applied to arbitrary invertible transfer functions ϕ. Indeed, the problem of learning associations $(G_i^\mu \rightarrow P^\mu)$ in a perceptron with an arbitrary invertible transfer function ϕ is equivalent to the problem of learning $(G_i^\mu \rightarrow \phi^{-1}(P^\mu))$ in a linear perceptron. All the results derived in this paper can then be applied to a perceptron with transfer function ϕ, except that μ_P and σ_P are now defined to be the two first moments of $\phi^{-1}(P_t^\mu)$.

Storage capacity

In the large N limit the probability of finding a set of weights that satisfies $P^\mu = P_t^\mu$ for all $\mu=1,\ldots,p$ is expected to be 1 if $\alpha \equiv p/N$ is below a critical value α_c, while it is 0 when $\alpha > \alpha_c$ [15]. α_c is therefore the number of associations that can be learned per synapse, and is commonly used as a measure of storage capacity.

This storage capacity can be computed analytically using the replica method (see Methods) [6,16,17,10,15]. The capacity is given by

$$\alpha_c = \int_B^\infty \frac{dt}{\sqrt{2\pi}} \exp(-t^2/2) \equiv H(B). \quad (2)$$

B is given by the equation

$$\frac{B}{G(B) - BH(B)} = \gamma, \quad (3)$$

$G(B) = \exp(-B^2/2)/\sqrt{2\pi}$, $H(B) = \frac{1}{2}(1 - \mathrm{erf}(B/\sqrt{2}))$, and γ depends on the statistics of the associations as

$$\gamma = \frac{\sigma_P^2 \mu_G^2}{\theta^2 \sigma_G^2}. \quad (4)$$

Therefore, the maximal capacity only depends on a single parameter γ, which is a function of the statistics of the patterns that need to be learned. This dependence is shown in Fig. 1A. It shows that the capacity is exactly equal to 0.5 when $\gamma=0$, while it decreases monotonically as γ increases.

If the number of patterns to be learned exceeds the maximal capacity, the mean squared error becomes strictly positive. It can also be computed using the replica method (see Methods, Eq. (17)). Unsurprisingly, it increases monotonically with α, as shown in Fig. 1B which shows the result of the analytical calculation, as well as numerical simulations. If uncorrelated noise is added to the perceptron, the total mean squared error is the sum of the error without noise (Eq. 17) and the variance of the uncorrelated noise.

In the simulations, inputs and outputs are drawn from an exponential distribution. The weight update at each presentation is the standard perceptron one, i.e.

$$\Delta w_i = \beta G_i (P_t - P), \quad (5)$$

where β is the learning rate. w_i is set to zero if application of the update leads to a negative weight. This corresponds to a gradient descent of a cost function proportional to $\sum_\mu (P^\mu - P_t^\mu)^2$, in the closed orthant $\{w_i \geq 0\}, i = \ldots, N$.

This learning rule is in qualitative agreement with experimental data on synaptic plasticity in GC to PC synapses [18,19]. In Purkinje cells, the error signal is thought to be conveyed by climbing fiber (CF) activation. Two protocols have been shown to be effective in eliciting long-term plasticity. Pairing GC with and CF activation leads to Long-Term Depression (LTD) of the synapse, while Long-Term Potentiation (LTP) is induced by stimulating the GC alone (see Fig. 3AB of [19] for details). Writing climbing fiber activation as $C = P - P_t + C_0$, we see that Eq. (5) is recovered if one chooses $\Delta w_i \propto G_i(C_0 - C)$, which captures the two experimental protocols described above.

Distribution of synaptic weights

The distribution of synaptic weights at maximal capacity can also be computed using the replica method (see [10] for details of the calculation). It turns out that the distribution obeys exactly the same equation as in the binary perceptron, i.e.

$$P(w) = H(-B)\delta(w) + \frac{1}{\sqrt{2\pi} w_s} \exp\left[-\frac{1}{2w_s^2} (w + B w_s)^2 \right] \Theta(w), \quad (6)$$

where

$$w_s = \frac{\overline{w}}{G(B) - BH(B)}, \quad (7)$$

and $\overline{w}$ is the average synaptic weight. In particular the fraction of zero weight synapses is $S = H(-B)$. Interestingly, there is a very simple relationship between capacity and fraction of silent synapses, $S + \alpha_c = 1$, that holds for any value of γ. The fraction of silent synapses S is shown as a function of γ in Fig. 2A. It shows that $S = 0.5$ when $\gamma = 0$, and increases monotonically with γ.

The full distribution of weights is shown in Fig. 2B, together with the results of a numerical simulation (see parameters in the caption of Fig. 2B). The theoretical distribution of synaptic weights is in good agreement with experimental measurements of the efficacy of a large set of GC to PC synapses, using paired recordings in vitro (see Fig. 6A of [10] for details) [20,21,10].

Above maximal capacity, $\alpha > \alpha_c$, the distribution of synaptic weights is still given by Eq. (6), but the fraction of zero weight synapses decreases monotonically with α, and goes to zero in the large α limit (see Fig. 2C). In that limit the distribution becomes

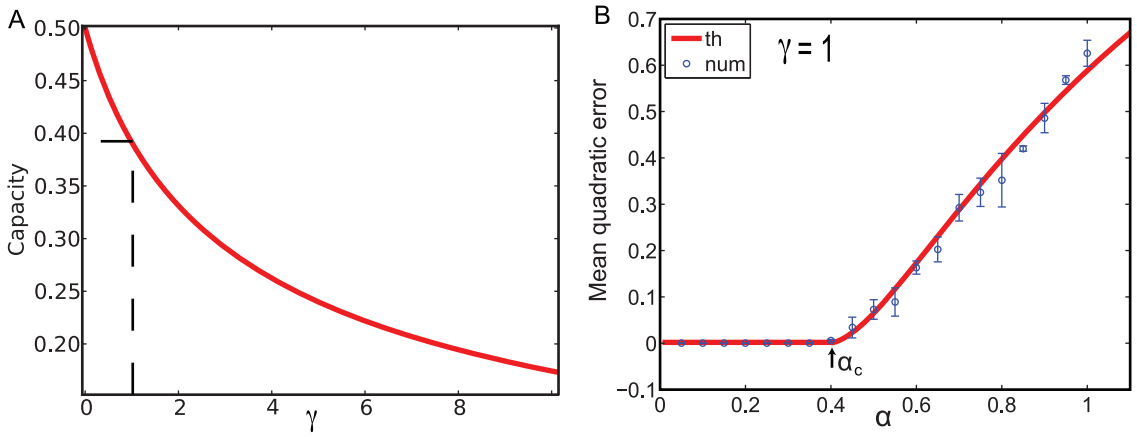

Figure 1. A. Maximal capacity as a function of γ. B. Mean squared error between the output P and the target output P_t as a function of α, for $\gamma=1$ ($\alpha_c\sim0.4$). Red: analytical calculation, Eq. (17); Blue, numerical simulations (with parameters: $N=1000$, $\theta=\sigma_P=\mu_P=\sigma_G=\mu_G=1$, $\beta=0.01$, simulation length $=100000N$, average over 20 trails, error bars: standard deviation).
doi:10.1371/journal.pcbi.1002919.g001

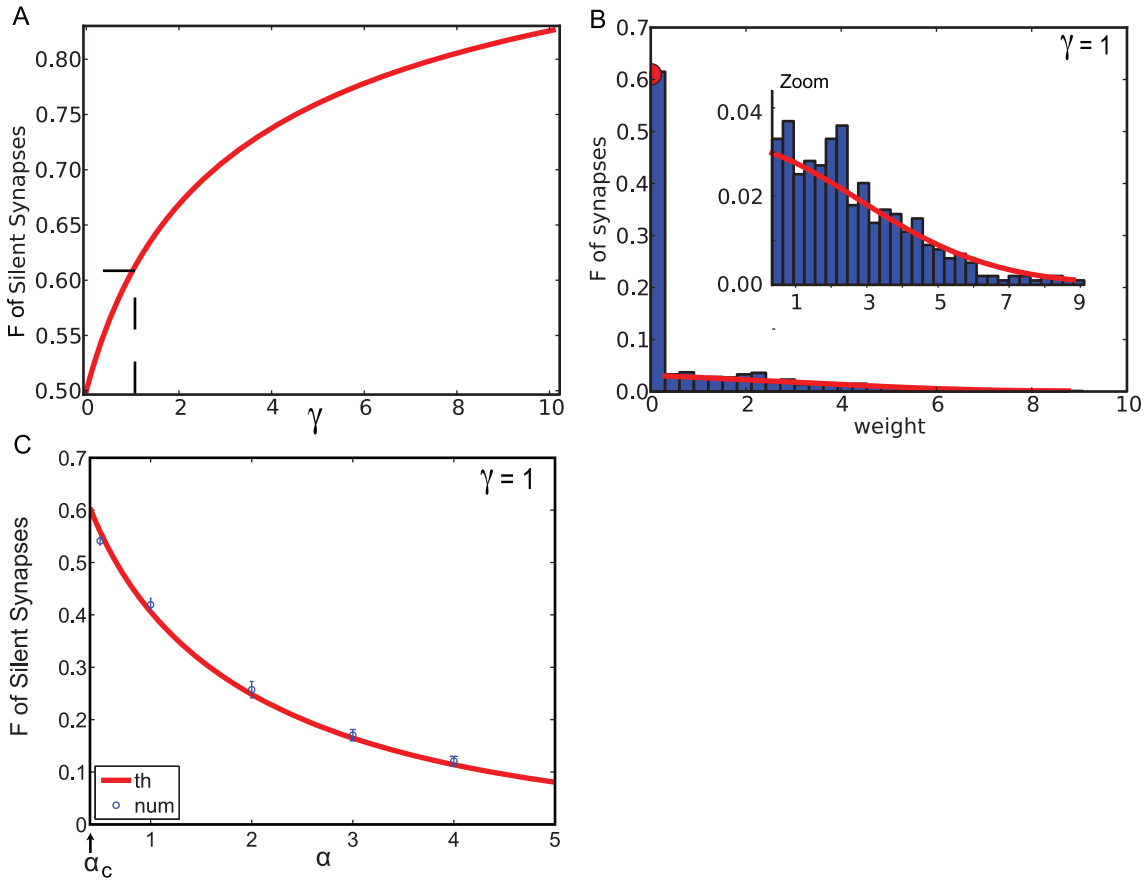

Figure 2. A. Fraction of silent synapses at maximal capacity as a function of γ. B. Distribution of synaptic weights for $\gamma=1$, at maximal capacity ($\alpha_c\sim0.4$). Red: analytical calculation, Eq. (6); Blue, numerical simulations (with parameters: $N=1000$, $\theta=\sigma_P=\mu_P=\sigma_G=\mu_G=1$, $\beta=0.01$, simulation length $=10000N$). C. Fraction of silent synapses as a function of α, beyond the maximal capacity ($\alpha_c\sim0.4$), for $\gamma=1$ (red: analytical calculation, $S=H(-B)$); blue: numerical simulations, with parameters $N=1000$, $\theta=\sigma_P=\mu_P=\sigma_G=\mu_G=1$, $\beta=0.01$, simulation length $=300000N$, average over 10 trails, error bars: standard deviation).
doi:10.1371/journal.pcbi.1002919.g002

increasingly close to a Gaussian distribution peaked around a positive value, with a width that tends to zero in the large α limit.

Statistics of inputs and outputs maximizing storage capacity

To maximize storage capacity, γ should be as small as possible. We first ask which distribution of inputs maximize capacity. From Eq. (4), it is clear that to maximize capacity, μ_G should be as small as possible, while σ_G should be as large as possible. Since ρ_{in} is a distribution of firing rates, it must be bounded between 0 and a maximal firing rate G_{max}. The distribution of a bounded variable that maximizes the variance with a fixed mean μ_G is a binary distribution $\rho_{in}(G)=(1-\mu_G/G_{max})\delta(G)+\mu_G/G_{max}\delta(G-G_{max})$. Thus, we predict that to optimize capacity, patterns of activity in the Granule cell layer should be sparse (to ensure μ_G is small), but active cells should be active close to their maximal firing rates. Interestingly, this is in striking agreement with available data [22,19,23] showing that (i) Granule cells have very sparse activity in vivo (average firing rates of 0.5 Hz [22]) (ii) they can respond with brief, high frequency bursts of action potentials to sensory inputs (with an average frequency of 77 Hz within the burst, and maximal frequencies up to 250 Hz, see e.g. Fig. 3 of [22]).

The next question is which distribution of output firing rates optimizes the capacity. Eq. (4) makes it clear the capacity is optimized for $\sigma_P=0$. In this limit however, all input patterns lead to exactly the same output, and the Purkinje cell output contains no information on which input was presented. This is of course not a desirable outcome, and suggests the capacity is not the correct measure to maximize in this case. We therefore turn to the Shannon mutual information between the Purkinje cell output and its inputs as a more suitable measure. In the presence of additive Gaussian noise of zero mean and standard deviation σ_n, this is simply the mutual information of a Gaussian channel with a signal-to-noise ratio σ_P^2/σ_n^2, i.e. $\log_2(1+\sigma_P^2/\sigma_n^2)/2$ bits per pattern (see e.g. [24]). The total information in bits per synapse is therefore $I=\alpha_c\log_2(1+\sigma_P^2/\sigma_n^2)/2$. The information is zero when $\sigma_P=0$, and reaches a maximum for a finite value of σ_P, which depends on both the noise standard deviation σ_n and $\sigma_{eff}=\sigma_G\theta/\mu_G$. Fig. 3A shows the information as a function of σ_P, for different values of σ_{eff}, for $\sigma_n=1$. It shows that the optimal value of σ_P increases approximately linearly with σ_{eff} for large σ_{eff} (see Fig. 3B).

Discussion

In this paper, we have considered an analog firing rate model for a Purkinje cell with plastic excitatory weights, and derived both its maximal capacity and the distribution of weights at maximal capacity. We showed that the capacity is of the same order as in a binary perceptron model.

The distribution of synaptic weights of the analog perceptron is composed at maximal capacity of two parts: a large fraction (>0.5) of silent synapses and a truncated Gaussian. It has exactly the same shape as in several other models: a standard binary perceptron [10], and a bistable perceptron [25]. This distribution is in quantitative agreement with a combination of electron microscopy and electrophysiological data in adult rat slices [20,21,10]. Furthermore, a gradient descent learning rule leading to maximal capacity bears strong similarities with synaptic plasticity experiments: LTD when PF and CF are coactivated, LTP when PF fires alone (i.e. CF below baseline, thus $P_t>P$) [18,19].

We found that in order to maximize the capacity, the input variance should be as large as possible. We argue that GCs in vivo are close to such an optimal distribution, since they fire high-frequency bursts at very low rates [22,19,23]. Furthermore, GC bursts have been found in some experiments to be critical to induce plasticity in PF to PC synapse [26]. Indeed, no plasticity is induced in those protocols with a single GC spike. Secondly, lower variance in the output also increases the capacity, but at a cost of losing information contained in the output, in the presence of noise. For a given variance of the noise, there is an optimal variance of the output that maximizes the information contained in the output.

The model we have studied here is essentially equivalent to the ADALINE (Adaptive Linear Neuron) model [27], whose storage capacity, in the absence of constraints on synaptic weights, is equal to 1. The result can be easily intuitively understood by the fact that when $\alpha=1$, there are exactly N linear equations to solve, Eq. 1, with N unknowns, w_i (see e.g. [15]). We have shown here that the constraints that all synaptic weights should be positive or zero leads to a capacity which is decreased by a factor 2 or more, depending on the value of γ. This decrease in capacity is similar to what is observed in the standard perceptron with excitatory synapses [7,8,9,10]. Note that learning associations with constrained weights is similar conceptually to non-negative matrix factorization [28,29]. Generalizations of such models in the temporal domain (the so-called adaptive filter models) have been proposed to describe learning in the cerebellar cortex [30,31,32,33]. It would be of interest to investigate capacity and distribution of synaptic weights of such models.

In this paper, we have focused on a single plasticity site, the GC to PC synapse. Many other sites of plasticity are known to exist in the cerebellum [18]. Future studies are needed to clarify the impact of these additional sites of plasticity on the learning capabilities of this structure.

Methods

Calculation of the storage capacity

The replica method involves calculating the average logarithm of the volume of the space of weights satisfying all constraints given by Eq. (1) [6]. To compute the average logarithm, one uses the replica trick: n replicas of the system are introduced, one computes

$$\langle V^n\rangle=\langle\int\prod_{i,a}dw_{ia}\prod_{\mu,a}\delta\left(\phi\left(\frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N}w_{ia}G_i^\mu-\theta N\right)\right)-P_t^\mu\right)\rangle,$$

where $\langle.\rangle$ represents an average over the patterns, and a is a replica index. This calculation is done using a standard procedure. After introducing integral representations for the delta functions, one averages over the distribution of the patterns. One then introduces order parameters

$$\frac{1}{N}\sum_j w_j^a=\frac{\theta}{\mu_G}+\frac{M^a}{\sqrt{N}}\equiv\overline{w}+\frac{M^a}{\sqrt{N}} \tag{8}$$

$$\frac{1}{N}\sum_j\left(w_j^a\right)^2=Q^a \tag{9}$$

$$\frac{1}{N}\sum_j w_j^a w_j^b=q^{ab}, \tag{10}$$

together with conjugate parameters $\hat{M}^a$, $\hat{Q}^a$ and $\hat{q}^{ab}$. We then use a replica-symmetric ansatz (all the order parameters are taken to

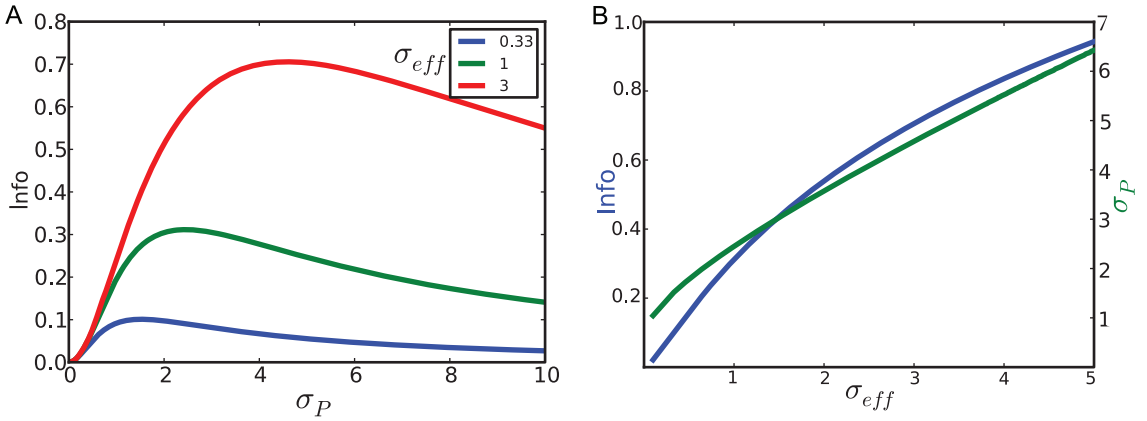

Figure 3. A. Information as a function of σ_P for different levels of $\sigma_{eff}=\sigma_G\theta/\mu_G$, and $\sigma_n=1$. B. Optimal σ_P (green line, right y-axis) and information (blue line, left y-axis) as a function of σ_{eff}.
doi:10.1371/journal.pcbi.1002919.g003

be independent of replica index a), perform the limit $n\to 0$ and obtain

$$\langle V^n\rangle \propto \int dMdQdqd\hat{M}d\hat{Q}d\hat{q}\exp(NnF), \tag{11}$$

$$\begin{aligned} F=&-\hat{Q}Q+\frac{1}{2}\hat{q}q+\overline{w}\hat{M}\\ &+\int_{-\infty}^{+\infty}\frac{du}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)\log\int_0^{\infty}dw\exp\left[(\hat{Q}-\frac{\hat{q}}{2})w^2+w(u\sqrt{\hat{q}}-\hat{M})\right]\\ &+\alpha\left(-\log(Q-q)-\frac{q}{2(Q-q)}-\frac{(\mu_G M-\mu_P)^2+\sigma_P^2}{2\sigma_G^2(Q-q)}\right), \end{aligned} \tag{12}$$

where in the Equation for F (Eq. (12)), the two first lines are identical to the binary perceptron with excitatory weights [10], while the last line is specific to the analog perceptron.

In the large N limit, the integral in Eq. (11) is dominated by the extremum of F. The typical values of all order parameters are then obtained by the resulting saddle point equations, setting the derivatives of F with respect to all order parameters to zero. The maximal capacity α_C is obtained in the limit $q\to Q$, for which the volume vanishes. This limit yields Eqs. (2,4).

Calculation of the mean squared error

Following [34], we introduce a cost function which is given by the sum of the squared error for all patterns,

$$C(w_i)=\sum_{\mu}\left(\frac{1}{\sqrt{N}}\left(\sum_i w_i G_i^{\mu}-\theta N\right)-P^{\mu}\right)^2 \tag{13}$$

and compute its minimum over the space of weights. This is done introducing a partition function $Z(h)$,

$$Z(h)=\int\prod_{i,a}dw_i\exp(-hC(w_i)) \tag{14}$$

where h is an inverse temperature, and computing $\langle\log Z(h)\rangle$ using the replica method. The mean squared error is then given by

$$E_{min}=\lim_{h\to\infty}-\frac{d}{dh}\frac{\langle\log Z(h)\rangle}{p} \tag{15}$$

To perform this calculation, a new parameter has to be introduced,

$$\rho=2h\sigma_G^2(Q-q) \tag{16}$$

which will remain finite when $\alpha>\alpha_c$ in the limit $h\to\infty$, $q\to Q$. The mean squared error is then given by

$$E_{min}=\frac{\sigma_G^2\overline{w}^2}{1+\rho}\left(\gamma+\frac{\rho K}{1+\rho}-2B\sqrt{\frac{K}{\alpha}}-\frac{1}{\alpha}\left((B^2+1)H(B)-BG(B)\right)\right) \tag{17}$$

where

$$K=\gamma+\frac{(1+B^2)H(B)-BG(B)}{(G(B)-BH(B))^2} \tag{18}$$

$$\alpha=(1+B^2)H(B)-BG(B)+\gamma(G(B)-BH(B))^2 \tag{19}$$

$$\rho=\frac{H(B)}{\alpha-H(B)} \tag{20}$$

When $\alpha=\alpha_c$, ρ diverges to infinity, $E_{min}=0$, and Eqs. (19,20) reduce to Eqs. (2,3).

Acknowledgments

We would like to thank Boris Barbour, Mariano Casado, Vincent Hakim, Clément Léna, and Jean-Pierre Nadal for fruitful discussions and helpful comments on the manuscript.

Author Contributions

Conceived and designed the experiments: NB. Performed the experiments: CC NB. Analyzed the data: CC NB. Contributed reagents/materials/analysis tools: CC NB. Wrote the paper: CC NB.

References

1. Marr D (1969) A theory of cerebellar cortex. J Physiol (Lond) 202: 437–470.
2. Albus J (1971) A theory of cerebellar function. J Mathematical Biosciences 10: 25–61.
3. Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psych Review 65: 386–408.
4. Soetedjo R, Kojima Y, Fuchs AF (2008) Complex spike activity in the oculomotor vermis of the cerebellum: a vectorial error signal for saccade motor learning? J Neurophysiol 100: 1949–1966.
5. Cover T (1965) Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans Electron Comput 14: 326.
6. Gardner E (1988) The phase space of interactions in neural network models. J Phys A 21: 257–270.
7. Amit D, Wong K, Campbell C (1989) Perceptron learning with sign-constrained weights. Journal of Physics A: Mathematical and General 22: 2039–2045.
8. Kanter I, Eisenstein E (1990) On the capacity per synapse. J Phys A: Math Gen 23: L93i.
9. Nadal JP (1990) On the storage capacity with sign-constrained synaptic couplings. Network: Comput Neural Syst: 463–466.
10. Brunel N, Hakim V, Isope P, Nadal JP, Barbour B (2004) Optimal information storage and the distribution of synaptic weights: Perceptron versus purkinje cell. Neuron 43: 745–757.
11. Barmack NH, Yakhnitsa V (2008) Functions of interneurons in mouse cerebellum. J Neurosci 28: 1140–1152.
12. Ke MC, Guo CC, Raymond JL (2009) Elimination of climbing fiber instructive signals during motor learning. Nat Neurosci 12: 1171–1179.
13. Thier P, Dicke PW, Haas R, Barash S (2000) Encoding of movement time by populations of cerebellar Purkinje cells. Nature 405: 72–76.
14. Thach WT (1968) Discharge of Purkinje and cerebellar nuclear neurons during rapidly alternating arm movements in the monkey. J Neurophysiol 31: 785–797.
15. Hertz J, Krogh A, Palmer RG (1991) Introduction to the Theory of Neural Computation. Redwood City CA: Addison-Wesley.
16. Gutfreund H, Stein Y (2613–2630) Capacity of neural networks with discrete synaptic couplings. Journal of Physics A: Mathematical and General 23: 1990.
17. Kohler H, Widmaier D (1991) Sign-constrained linear learning and diluting in neural networks. Journal of Physics A: Mathematical and General 24: L495–L502.
18. Hansel C, Linden DJ, D'Angelo E (2001) Beyond parallel fiber LTD: the diversity of synaptic and non-synaptic plasticity in the cerebellum. Nat Neurosci 4: 467–475.
19. Jorntell H, Hansel C (2006) Synaptic memories upside down: bidirectional plasticity at cerebellar parallel fiber-Purkinje cell synapses. Neuron 52: 227–238.
20. Harvey RJ, Napper RM (1988) Quantitative study of granule and Purkinje cells in the cerebellar cortex of the rat. J Comp Neurol 274: 151–157.
21. Isope P, Barbour B (2002) Properties of unitary Granule cell to Purkinje cell synapses in adult rat cerebellar slices. J Neurosci 22: 9668–9678.
22. Chadderton P, Margrie TW, Hausser M (2004) Integration of quanta in cerebellar granule cells during sensory processing. Nature 428: 856–860.
23. Rancz EA, Ishikawa T, Duguid I, Chadderton P, Mahon S, et al. (2007) High-fidelity transmission of sensory information by single cerebellar mossy fibre boutons. Nature 450: 1245–1248.
24. Cover T, Thomas J (1991) Elements of Information Theory. New York: Wiley.
25. Clopath C, Nadal JP, Brunel N (2012) Storage of correlated patterns in standard and bistable purkinje cell models. Plos Comp Biol 8: e1002448.
26. Bidoret C, Ayon A, Barbour B, Casado M (2009) Presynaptic NR2A-containing NMDA receptors implement a high-pass filter synaptic plasticity rule. Proc Natl Acad Sci USA 106: 14126–14131.
27. Widrow B, Hoff ME (1960) Adaptive switching circuits. In: 1960 IRE WESCON Convention Record. New York: IRE. pp. 96–104.
28. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401: 788–791.
29. Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorizatio. Adv Neural Info Proc Syst 13: 556–562.
30. Fujita M (1982) Simulation of adaptive modification of the vestibulo-ocular reflex with an adaptive filter model of the cerebellum. Biol Cybern 45: 207–214.
31. Dean P, Porrill J, Stone JV (2002) Decorrelation control by the cerebellum achieves oculomotor plant compensation in simulated vestibulo-ocular reflex. Proc Biol Sci 269: 1895–1904.
32. Porrill J, Dean P (2007) Cerebellar motor learning: when is cortical plasticity not enough? PLoS Comput Biol 3: 1935–1950.
33. Lepora NF, Porrill J, Yeo CH, Dean P (2010) Sensory prediction or motor control? Application of marr-albus type models of cerebellar function to classical conditioning. Front Comput Neurosci 4: 140.
34. Gardner E, Derrida B (1988) Optimal storage properties of neural network models. J Phys A:Gen 21: 271–284.