



HAL
open science

NEURAL NETWORKS : A STATISTICAL PHYSICS APPROACH

J.-P. Nadal

► **To cite this version:**

J.-P. Nadal. NEURAL NETWORKS : A STATISTICAL PHYSICS APPROACH. Journal de Physique Colloques, 1989, 50 (C3), pp.C3-223-C3-227. 10.1051/jphyscol:1989334 . jpa-00229475

HAL Id: jpa-00229475

<https://hal.science/jpa-00229475>

Submitted on 4 Feb 2008

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.

NEURAL NETWORKS : A STATISTICAL PHYSICS APPROACH

J.-P. NADAL

*Laboratoire de Physique Statistique de l'Ecole Normale Supérieure, 24,
rue Lhomond, F-75231 Paris Cedex 05, France*

Résumé. Cette contribution présente le formalisme et quelques résultats importants spécifiques de l'étude des réseaux de neurones avec les outils de la physique statistique.

Abstract. This contribution introduces the formalism and some important results specific of the statistical physics approach to neural network modeling.

1. BASICS

A general introduction to neural network modeling has been given in this workshop by G. Dreyfus. Other reviews can be found in /1,2,3,4/. I will present here some aspects of the specific approach of people working in statistical physics. In this section I will introduce the formalism and the main results on the Hopfield model of neural network. In section 2 and 3 I will present recent results respectively on the storage capacity and on the dynamics of neural networks.

We are interested in the properties of a large number N of formal neurons (the large N limit being the analog of the thermodynamic limit). Each formal neuron i , $i=1,N$, is described by a spin like variable, S_i , being $+1$ (neuron firing) or -1 (neuron quiescent). The dynamics of this network of N neurons is governed by the synaptic efficacies J_{ij} which characterize the property of the synapse from neuron j to neuron i . Considering the high connectivity in the cortex, a simple choice is to connect each neuron to every other neuron : $J_{ij} \neq 0$ for all $(i \neq j)$. Each neuron compute its "local field" $h_i(t)$:

$$h_i(t) = \sum_j J_{ij} S_j(t) \quad (1)$$

and

$$S_i(t + \Delta t) = \begin{cases} +1 & \text{with probability } 1/[1 + \exp - 2\beta h_i(t)] \\ -1 & \text{with probability } 1/[1 + \exp + 2\beta h_i(t)] \end{cases} \quad (2)$$

where $\beta = 1/T$ is a temperature-like parameter modeling the intrinsic noise of the system. For random sequential updating, the dynamics corresponds to the usual Monte Carlo algorithm. Hopfield /5/ pointed out that if we make

the (unrealistic) assumption that the J_{ij} are symmetric, then one can define an energy E :

$$E = -\frac{1}{2} \sum_{i \neq j} J_{ij} S_i S_j \quad (3)$$

and we end up with a spin like system.

Now we would like to choose the J_{ij} such that a set $\{\xi_i^\mu, i = 1, N\}, \mu=1, p$ of p patterns to be learned are fixed points of the dynamics. The standard choice/5/ follows the "generalized Hebb rule" :

$$i \neq j \quad J_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \quad (4)$$

Different versions of this model have been introduced and studied by many authors /4,5,6/. Two points were made clear by the elegant presentation of Hopfield /5/: first, recognition of a stored pattern is viewed as the convergence of a dynamical system towards an attractor - the set of activity coding this pattern. This notion of memories as attractors renders natural the possibility of retrieving a complete pattern starting with a noisy or incomplete pattern. Second, as shown above, the formalism make explicit the link with statistical physics, so that it is possible to apply the methods of this domain.

In the framework of statistical physics, it is natural to work with an ensemble of patterns with the same statistical properties. In particular, if we choose the ξ^μ at random, with $\xi_i^\mu = \pm 1$ with equal probability, J_{ij} can be either positive (the synapse is then excitatory) or negative (the synapse is inhibitory): indeed we know from the spin-glass theory /7/ that such a situation can lead to many (meta)stable states. Noting that (3) can be rewritten

$$E = -\frac{1}{2} \frac{1}{N} \sum_{\mu} \left(\sum_i \xi_i^\mu S_i \right)^2 + \text{constant}$$

we see that if the patterns are orthogonal, the ground states of the system are exactly the patterns ξ^μ . For random patterns, the patterns are only statistically orthogonal. In this case however, the thermodynamics of the model can be solved /2/. The main results are the following: if the number of pattern p is smaller than a critical value $p_c = \alpha_c(T)N$, with $\alpha_c \sim 0.14$ at $T=0$, then the network operates as an associative memory. Starting with an initial configuration not too far from a stored pattern ξ^μ , the network converges toward a fixed point S^μ which is very close to the stored pattern if we call m_μ the overlap of S^μ on ξ^μ :

$$m_\mu = \left\langle \frac{1}{N} \sum_i S_i^\mu \xi_i^\mu \right\rangle \quad (5)$$

m_μ is at least equal to 0.97 (at zero temperature). If $\alpha = p/N$ is greater than α_c , the system enters a spin-glass phase, and all retrieval properties are lost.

Many studies have been made on variants of this model - e.g. in order to store patterns of a given mean activity level/2,8/. Simple modifications of

the learning rule can prevent from the deterioration of the memory: one obtains short-term memory models, where new patterns can always be learned whereas older patterns are progressively forgotten /9/. It has been shown that random dissymmetrization or dilution of the network do not affect qualitatively the properties of the system/2/. Moreover the dynamics of a highly diluted and asymmetric model/10/ can be solved exactly. The main qualitative results are that one keeps the notion of memories as attractors, and that the capacity is proportional to the connectivity C - that is the typical number of neurons to which any neuron is connected :

$$p_c = \alpha_c C . \quad (6)$$

2. MAXIMAL STORAGE CAPACITY

The scaling (6) is not surprising: the information concerning the p patterns of N "bits" is stored in NC synaptic efficacies, so that we can expect $pN \sim NC$. One can then ask: what is the maximal possible capacity (for some optimized choice of the J_{ij})? For random patterns the nonintuitive result is/11,12/:

$$\alpha_{\max} = 2 \quad (7)$$

Note that this is the optimal capacity when one requires *exact* retrieval ($m_\mu = 1$ for every μ). The next question is of course: is there any algorithm which allows to reach this theoretical optimum? The answer is yes: there are iterative learning algorithms/12,13,14/, all of them being variant of the 'Perceptron algorithm'/15/(introduced 20 years ago!). This algorithm is simple. First, note that a pattern ξ^μ is a fixed point of the dynamics (at $T=0$) if

$$\text{for all } i, i=1, N \quad \xi_i^\mu h_i^\mu > 0 \quad (7)$$

where h_i^μ is the local field when in the state ξ^μ :

$$h_i^\mu = \sum_j J_{ij} \xi_j^\mu \quad (8)$$

For each site i ($i=1, N$), the algorithm compute (J_{ij} , $j=1, N$). It proceeds as follows, for each value of i :

$t=0$. Start with $J_{ij} = 0$, $j=1, N$

* Pick a pattern μ at random

If $\xi_i^\mu h_i^\mu > 0$, go to *

else, $t \leftarrow t+1$; ($J_{ij} \leftarrow J_{ij} + \xi_i^\mu \xi_j^\mu$), $j=1, N$; go to *

A theorem/15/ states that, if a solution exists (that is if there exists at least one matrix such that (7) is true for every pattern), then after a finite number of steps t , the algorithm gives a solution.

3. STABILITY AND BASINS OF ATTRACTION

So far nothing has been said on the associative properties of the network: once a set of patterns has been learned with a specific learning rule, one would like to know what are the basins of attraction associated

with each pattern. Two important results have been obtained recently. First, the basins of attraction are circular in probability: an initial configuration $S_{j,i=1,N}$ with an overlap q with one of the stored patterns, will evolve with probability one towards this pattern (under the dynamics (1),(2)) if q is greater than some critical value q_c /16,17/. Second, the value of q_c depends mainly on one family of parameters, called the stabilities, defined for each pattern by

$$\Delta_{i\mu} = \xi_{i\mu} \sum_j J_{ij} \xi_{j\mu} / [\sum_j J_{ij}^2]^{1/2} \quad (9)$$

(note the normalization which fix the scale of the coupling matrix). As seen above, a pattern is stable if all these stabilities are positive. The result here is that the greater they are the larger the basin of attraction (the smaller q_c)/16-18/. The stabilities are the main parameters controlling the dynamics. The next important parameter is the degree of symmetry of the matrix J_{ij} : knowing the stabilities and the degree of symmetry of the matrix (J_{ij}), one can compute a good approximation of q_c /17/.

Thus if one wants to have the best possible associative properties one would like a learning rule which gives large values of stabilities. In fact an algorithm, again a variant of the Perceptron algorithm, has been proposed recently /14/ which allows to reach the largest possible stability K , where K is the smallest of all the stability parameters. Given a set of patterns - any set, not necessarily random-, any Perceptron type algorithm will find one solution - provided there exists at least one. The optimal algorithm will find the solution which gives the best possible minimal stability for this set of patterns, insuring thus the largest possible basin of attraction.

CONCLUSION

I have presented some of the recent results obtained within the framework of statistical physics. These results show the progress made in the comprehension in the associative properties of neural networks using a typical strategy of statistical physics: models are analyzed in the large N limit, and quantitative results are obtained by considering statistical ensembles (here of patterns).

ACKNOWLEDGEMENTS

I had the pleasure to work with W. Krauth and M. Mézard in contributing to some of the recent results presented here.

REFERENCES

- / 1 / Cowan J.D. and Sharp, D.H., preprint 1987; Tank D.W. and Hopfield, J.J., Scientific American **257**, 62 (1987); Bienenstock, E., Fogelman, F. and Weisbuch, G., "Disordered Systems and Biological Organization", (Springer-Verlag, Berlin 1986).
- / 2 / Amit, D.J., in Heidelberg Colloquium on Glassy Dynamics, von Hemmen, J.L. and Morgenstern, I., eds., (Springer-Verlag, Berlin 1987), p.430.

- / 3 / Kohonen, T., "Self Organization and Associative Memory", (Springer-Verlag, Berlin 1984).
- / 4 / Rumelhart, D.E., Hinton G.E. and Williams, R.J., Nature 323, 533 (1986).
- / 5 / Hopfield, J.J., Proc. Natl. Acad. Sci. USA 79, 2554 (1982).
- / 6 / See references in Shinomoto, S., Biol. System 57, 197 (1987) and in Cowan and Sharp /1/.
- / 7 / Mézard, M., Parisi, G. and Virasoro, M.A., "Spin-Glass Theory and Beyond" (World Scientific, 1988).
- / 8 / Parga, M. and Virosoro, M.A., J. Physique (Paris) 47, 1857 (1986); Feigelman, M.V. and Ioffe, L.B., Int. J. of Mod. Phys. B1, 51 (1987); Gutfreund, H., preprint 1987; Cortes, C., Krogh, A. and Hertz, J.A., J. Phys. A.
- / 9 / Nadal, J.P., Toulouse, G., Changeux, J.P. and Dehaene, S., Europhys. Lett. 1, 535 (1986); Parisi, G., J. Phys. A19, L617 (1986); Vedenov, A.A. and Levchenko, E.B., JETP Lett. 41, 402 (1985); Mézard, M., Nadal, J.P., Toulouse, J. Physique (Paris) 47, 1457 (1986).
- / 10 / Derrida, B., Gardner, E. and Zippelius, A., Europhys. Lett. 4, 167 (1987).
- / 11 / Venkatesh, S., Proceedings of the Conference on Neural Networks for Computing (Snowbird, 1986); T.M. Cover, IEEE transactions EC14, 326 (1985).
- / 12 / Gardner, E., Europhys. Lett. 4, 481 (1987); J. Phys. A21, 257 (1988).
- / 13 / Diederich, S. and Oppen, M., Phys. Rev. Lett. 58, 949 (1987); Pöppel, G. and Krey, U., Europhys. Lett. 4, 979 (1987); Kleinfeld, D. and Pendergraft, D.B., Biophys. 51, 47 (1987)
- / 14 / Krauth, W. and Mézard, M., J. Phys. A20, L745 (1987).
- / 15 / Minsky, M. and Papert, S., "Perceptrons" (Cambridge, MIT Press, 1969).
- / 16 / Forrest, B.M., J. Phys. A: Math. Gen. 21, 245 (1988).
- / 17 / Krauth, W., Nadal, J.P. and Mézard, M., J. Phys. A: Math. Gen. 21, 2995 (1988).
- / 18 / Krauth, W., Mézard, M. and Nadal, J.P., to appear in Complex Syst.