

Classification

Physics Abstracts

87.10 — 64.40 — 75.10H

Storage of spatially correlated patterns in autoassociative memories

Rémi Monasson

Laboratoire de Physique Théorique de l'Ecole Normale Supérieure (*), 24 rue Lhomond,
75231 Paris Cedex 05, France

(Received 2 November 1992, accepted in final form 8 January 1993)

Abstract. — The effects of spatially organised data on autoassociative neural networks are investigated in the optimal storage case. An analytical study is possible for weak spatial correlations. It predicts an increasing of the storage capacity α_c and ferromagnetic means for the couplings. Numerical simulations confirm these results for large spatial correlations.

1. Introduction.

Since it was introduced, the concept of attractor neural network has gained much attention from the physics community [1]. Numerical and theoretical studies have already captured many interesting features (e.g. working as an associative memory, robustness against degradation, fast parallel processing, learning from examples, .. [2]). For mainly technical reasons, these results were obtained in the absence of any spatial structure of both the neural networks and the inputs. However most real life applications deal with spatially organised data.

In this paper, we focus on the effects of such a spatial organisation of the input patterns on the behaviour of an autoassociative neural network. Such a network exists in a d -dimensional physical space and realistic patterns exhibit spatial correlations.

The aim is twofold. First, it is of interest to see how the results derived for non spatially structured networks are modified in this new framework. We shall concentrate upon the optimal storage capacity, using the statistical mechanics tools developed by Gardner-Derrida [3]. Secondly, we shall investigate how the network organises itself (from the weights distribution standpoint) in the presence of spatially correlated patterns. It has already been shown for hetero-associative mapping [4] and information processing [5], [6] that the resulting couplings organisation is non trivial. We shall see here that the latter is deeply different from the usual uncorrelated case and influences the retrieval properties of the network near the saturation.

(*) Unité Propre de Recherche du CNRS, associée à l'Ecole Normale Supérieure et à l'Université de Paris-Sud

2. Description of the model.

The network consists of N neurons, each connected to every other, and taking a value ± 1 . Neuron S_i is connected to neuron S_j via a real valued synapse J_{ij} ($J_{ii} = 0, \forall i$). The dynamics of the system follows a sequential updating rule of the form

$$S_i(t+1) = \text{sign} \left(\sum_{j(\neq i)} J_{ij} S_j(t) \right) \quad (1)$$

The training set consists of $P = \alpha N$ binary patterns $\{\xi_i^\mu\}$ belonging to a d -dimensional space. Zero-mean activity and spatial correlation are imposed by the constraints

$$\overline{\xi_i^\mu} = 0 \quad \text{and} \quad \overline{\xi_i^\mu \xi_j^\sigma} = \delta^{\mu\sigma} C_{ij} \quad (2)$$

where the overbars represent an average over the probability distribution of the pattern bits. C is a symmetric and positive matrix with a 1 on the diagonal, dependent on the distance $|i - j|$ (the subscripts i and j may be seen as d -dimensional vectors but all the simulations will be performed here for $d = 1$). Binary patterns obviously can not be entirely defined by the two first moments of their distribution. We shall discuss this point in the next part.

The stability of the pattern μ at site i is given by

$$\Delta_i^\mu = \xi_i^\mu \sum_{j(\neq i)} J_{ij} \xi_j^\mu. \quad (3)$$

The training set is said to be stored if all the patterns are fixed points of the dynamics (1), i.e. if all the stabilities are positive. However, an associative memory network requires that large attraction basins surround these stored patterns. It has been shown (at least for uncorrelated data) that their size increases with the value of the parameter $\kappa = \min_{i,\mu} (\Delta_i^\mu)$ [8].

We further define a quantity which will appear in the following work. It is the $N - 1$ dimensional matrix

$$\tilde{C}_{ij} = C_{ij} - C_{i1} C_{j1} \quad (i, j = 2 \dots N) \quad (4)$$

One can easily verify that the inverse matrix is

$$\left(\tilde{C}^{-1} \right)_{ij} = \left(C^{-1} \right)_{ij} \quad (i, j = 2 \dots N) \quad (5)$$

We define the normalised trace of any matrix A by $\text{Tr}A = \frac{1}{N} \sum_i A_{ii}$.

3. Analytical study of the network.

In this part, we try to answer the following question : what is the maximal size α of the training set the above network is able to store ? Given a training set $\{\xi_i^\mu\}$, for each possible synaptic matrix $\{J_{ij}\}$, we define an error function

$$E(\{J_{ij}\}, \{\xi_i^\mu\}) = \sum_{\mu=1}^P \sum_{i=1}^N \theta(\kappa - \Delta_i^\mu) \quad (6)$$

where κ is a positive parameter. Using the framework created by Gardner-Derrida [3], the partition function of the network is

$$Z(\{\xi_i^\mu\}) = \int \prod_{i \neq j} dJ_{ij} \prod_i \delta \left(\sum_{j(\neq i)} J_{ij}^2 - 1 \right) \exp(-\beta E(\{J_{ij}\}, \{\xi_i^\mu\})) \quad (7)$$

where β plays the role of an inverse temperature. In the following, we shall send $\beta \rightarrow \infty$, so that the integration in (7) takes into account the synaptic weights storing all the patterns with stabilities larger than κ . The choice of the cost function (6) is natural in the zero temperature limit only below the critical capacity. Training the network with other energies E (including error-size, and not only error-number, dependences) may be more efficient at relatively small β and high storage levels to improve the retrieval performances of the network [9].

Using the replica method, we compute $\overline{\ln Z}$, which is assumed to be self-averaging in the limit of large network size. Since the domain of suitable couplings remains connected - even with correlated patterns - we make the replica symmetric Ansatz. This Ansatz has also been shown to provide a stable saddle-point for hetero-associative storage [4].

3.1 THE REPLICA CALCULATION OF THE PARTITION FUNCTION. — The network partition function Z defined in (7) is equal to the product of the N single site partition functions Z_i where $i = 1 \dots N$.

$$Z_i = \int \prod_{j(\neq i)} dJ_{ij} \delta \left(\sum_{j(\neq i)} J_{ij}^2 - 1 \right) \exp \left(-\beta \sum_{\mu=1}^P \theta \left(\kappa - \xi_i^\mu \sum_{j(\neq i)} J_{ij} \xi_j^\mu \right) \right) \quad (8)$$

and in the large N limit [3],

$$\frac{1}{N^2} \overline{\ln Z} = \frac{1}{N^2} \sum_{i=1}^N \overline{\ln Z_i} = \frac{1}{N} \overline{\ln Z_1} = \lim_{N \rightarrow \infty} \lim_{n \rightarrow 0} \frac{\overline{Z_1^n} - 1}{nN} \quad (9)$$

We introduce n replicas $\{J_{ij}^a\}$, $a = 1 \dots n$ and average over the patterns distribution (2)

$$\begin{aligned} \overline{Z_1^n} = & \int \prod_{j,a} dJ_{1j}^a \prod_a \delta \left(\sum_j (J_{1j}^a)^2 - 1 \right) \prod_{\mu,a} \left(\frac{dt_\mu^a}{2\pi} \right) \\ & \times \exp \left(-\beta \sum_{\mu,a} \theta(\kappa - t_\mu^a) + \sum_{\mu,a} \hat{t}_\mu^a t_\mu^a + \sum_\mu I_\mu \right) \end{aligned} \quad (10)$$

where

$$\begin{aligned} I_\mu = \ln \left[\overline{\exp \left(- \sum_{a,j} \hat{t}_\mu^a J_{1j}^a \xi_1^\mu \xi_j^\mu \right)} \right] &= - \left[\sum_a \hat{t}_\mu^a \sum_j J_{1j}^a C_{1j} \right] + \left[\sum_{ab} \hat{t}_\mu^a \hat{t}_\mu^b \sum_{jk} J_{1j}^a J_{1k}^b \tilde{C}_{jk} \right] \\ &- \left[\sum_{abc} \hat{t}_\mu^a \hat{t}_\mu^b \hat{t}_\mu^c \sum_{jkl} J_{1j}^a J_{1k}^b J_{1l}^c C_{jkl}^{(3)} \right] + \dots \end{aligned} \quad (11)$$

The tensor appearing in the third term of the r.h.s. equals

$$C_{jkl}^{(3)} = \overline{\xi_1^j \xi_j^k \xi_k^l} - C_{1j} C_{kl} - C_{1k} C_{jl} - C_{1l} C_{jk} + 2 C_{1j} C_{1k} C_{1l} \quad (12)$$

We may assume, as for the hetero-associative storage [4], that the patterns satisfy some clustering conditions, i.e. that the k -points connected correlation function of their probability distribution always decrease exponentially with the distance separating the two closest points among the k ones.

For any pattern μ , the products $\xi_1^\mu \xi_j^\mu$ are in average equal to C_{1j} . The stability of the pattern μ defined by formula (3) will be increased if the synapses between neighbouring neurons are strengthened with the correct signs such that $J_{1j} C_{1j} > 0, \forall j$. It is thus plausible that

$J_{1j} = O(1)$ when the neuron j is close to the neuron 1, while it is of order $\frac{1}{\sqrt{N}}$ in the usual uncorrelated case. As a consequence, the truncation of the I_μ 's at the quadratic term in \hat{t} is not valid, even in the large N limit (as it used to be in [4]). An exact calculation of $\overline{\ln Z}$ for any spatially correlated distribution of the input patterns does not seem possible. It would both require the knowledge of all the moments of the statistical distribution of the patterns and the computation of non Gaussian integrals in \hat{t} .

In the following, we resort to an expansion of the partition function around the usual uncorrelated case ($C_{ij} \simeq \delta_{ij}$), keeping only the first two terms of the expansion (11). To be more precise, when $C_{ij} = x^{|i-j|}$ and $x \ll 1$, this Gaussian approximation can be shown to provide the critical capacity up to x^2 .

3.2 THE RESULTS FOR WEAK CORRELATIONS. — For $C_{ij} \simeq \delta_{ij}$, the entropy of the synaptic weights is equal to

$$\begin{aligned} \frac{1}{N^2} \overline{\ln Z} \simeq & \frac{\ln 2\pi}{2} - \frac{1}{2} q \hat{q} + s \hat{s} + m \hat{m} + \hat{u} - \frac{1}{2} \text{Tr} \left[\frac{\hat{q} \tilde{C}}{2\hat{u}I + (2\hat{s} - \hat{q})\tilde{C}} \right] \\ & - \frac{1}{2} \text{Tr} \left[\ln(2\hat{u}I + (2\hat{s} - \hat{q})\tilde{C}) \right] \\ & + \frac{1}{2} \hat{m}^2 \sum_{j,k=2}^N C_{j1} [(2\hat{u}I + (2\hat{s} - \hat{q})\tilde{C})^{-1}]_{jk} C_{k1} + \alpha \int Dz \ln \left[H \left(\frac{z\sqrt{q-m+\kappa}}{\sqrt{s-q}} \right) \right] \end{aligned} \quad (13)$$

where

$$Dz = \frac{e^{-z^2/2}}{\sqrt{2\pi}}, \quad H(z) = \int_z^{+\infty} Dt \quad (14)$$

The three order parameters m, q, s are given by

$$\begin{aligned} m &= \sum_{j=2}^N C_{1j} \overline{\langle J_{1j} \rangle} \\ q &= \sum_{j,k=2}^N \tilde{C}_{jk} \overline{\langle J_{1j} \rangle \langle J_{1k} \rangle} \\ s &= \sum_{j,k=2}^N \tilde{C}_{jk} \overline{\langle J_{1j} J_{1k} \rangle} \end{aligned} \quad (15)$$

where $\langle . \rangle$ denotes the thermal average over the couplings J_{ij} . Variables $\hat{m}, \hat{q}, \hat{s}$ are Lagrange parameters enforcing constraints (15). \hat{u} ensures the normalisation of each line of the synaptic matrix to 1. These seven parameters are found by solving the saddle-point equations of (13).

We restrict our analysis to the critical line $\alpha_c(\kappa)$ where $s - q \rightarrow 0$ and the four Lagrange multipliers diverge. We then obtain two coupled equations linking the common critical value s_c of q, s to the critical parameter m_c

$$\begin{aligned} \kappa m_c &= \nu_c \sum_{j,k=2}^N C_{j1} [(\tilde{C} + \nu_c I)^{-1}]_{jk} C_{k1} \\ \frac{\text{Tr} \left[\frac{\tilde{C}}{(\tilde{C} + \nu_c I)^2} \right]}{\text{Tr} \left[\frac{\tilde{C}}{\tilde{C} + \nu_c I} \right]} &= \frac{\kappa^2 - \nu_c^2 \sum_{j,k=2}^N C_{j1} [(\tilde{C} + \nu_c I)^{-2}]_{jk} C_{k1}}{\kappa^2 (s_c + \nu_c) - \nu_c^2 \sum_{j,k=2}^N C_{j1} [(\tilde{C} + \nu_c I)^{-1}]_{jk} C_{k1}} \end{aligned} \quad (16)$$

where

$$\nu_c(\kappa, m_c, s_c) = \kappa (\kappa - m_c) + \kappa \sqrt{\frac{s_c}{2\pi}} \frac{e^{-\frac{(m_c - \kappa)^2}{2s_c}}}{H\left(\frac{m_c - \kappa}{\sqrt{s_c}}\right)} \tag{17}$$

Once m_c and s_c have been computed, the critical capacity $\alpha_c(\kappa)$ is given by

$$\alpha_c = \frac{1}{H\left(\frac{m_c - \kappa}{\sqrt{s_c}}\right)} \text{Tr} \left[\frac{\tilde{C}}{\tilde{C} + \nu_c I} \right] \tag{18}$$

Let us now apply these results to the one-dimensional case with exponentially decreasing correlations $C_{ij} = x^{|i-j|}$, where $x \ll 1$. The optimal capacity is reached when $\kappa = 0$ and from equations (16 - 18), we obtain

$$\alpha_c(x) = 2 + \frac{4}{\pi} x^2 + O(x^4) \tag{19}$$

Let us evaluate indeed the value of the first neglected term (that involving $C^{(3)}$, see (12)). As expected, the J 's do not vanish for large N and are order x or less when x is small. The cubic term in \hat{t} therefore scales as at least x^3 . This scaling holds for the higher terms in the expansion (11) while the quadratic term in \hat{t} is of order x^2 . We conclude that the results derived when forgetting all but the first two terms are valid up to order x^2 .

4. Interpretation and simulations.

4.1 THE STORAGE CAPACITY AND THE QUANTITY OF INFORMATION. — The results derived at the end of the previous part show that the critical capacity of attractor neural networks increases with the presence of spatial correlations in the training patterns. For feed-forward networks and hetero-associative mapping, the storage capacity remained equal to 2, whatever C_{ij} [4].

The patterns presenting the spatial correlations $C_{ij} = x^{|i-j|}$ may be generated as equilibrium configurations of a one dimensional Ising model at temperature T such that $x = \tanh(1/T)$. Thus the quantity of information $i(x)$ in a pattern is simply the entropy of this Ising lattice

$$i(x) = N \left[- \left(\frac{1-x}{2} \right) \ln \left(\frac{1-x}{2} \right) - \left(\frac{1+x}{2} \right) \ln \left(\frac{1+x}{2} \right) \right] \tag{20}$$

When $x \ll 1$, the total quantity of information $I(x)$ stored in the training set is

$$I(x) = \alpha_c(x) N \cdot i(x) = N^2 \left[2 \ln 2 + \left(\frac{4 \ln 2}{\pi} - 1 \right) x^2 + O(x^4) \right] \tag{21}$$

Though the capacity of auto-associative networks increases when they are presented with spatially correlated data, we see that (as far as we can trust the small x result) the quantity of information the neural network actually stores decreases. This situation already occurred with biased patterns [3].

Some numerical simulations were carried out to check the increase of the storage capacity with the amount of spatial correlations inside the training set. The patterns are generated through a Monte Carlo simulation of the Ising model and the Minover algorithm finds the couplings achieving the largest stability κ [4], [7]. The sizes of the network were equal to $N = 50, 100$ and 200 . The differences between these three simulations were lower than the error

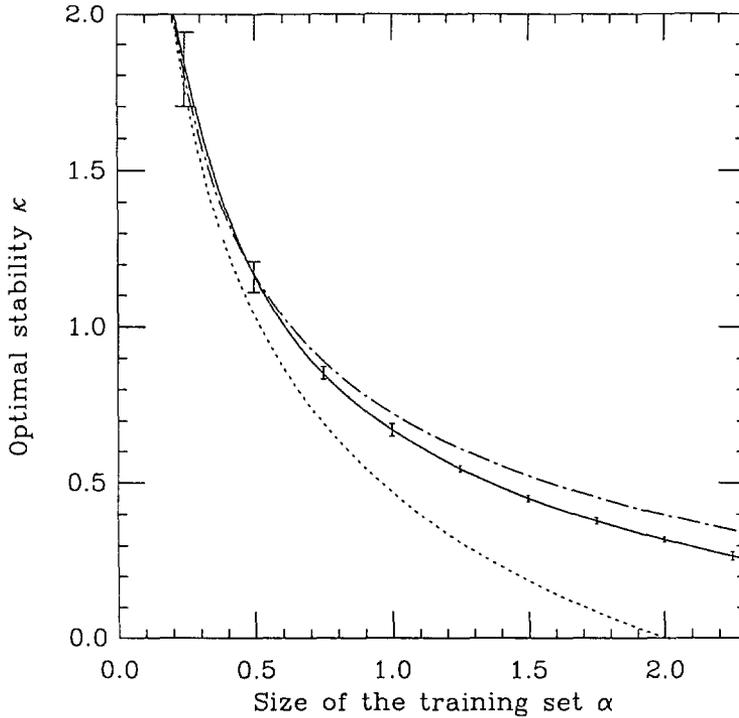


Fig.1. — the optimal stability κ for different sizes α obtained from the Minover algorithm (the full curve is a polynomial fit). The patterns are drawn from a one dimensional Ising model ($x = 0.7$ - see text) and the number of neurons is $N = 200$. The dotted line is the theoretical curve for the usual uncorrelated case ($x = 0$). The dash-dotted curve displays the solution of the saddle-point equations valid for low x .

bars (due to the statistical fluctuations and the finite running time of the Minover). We plot here the results obtained for $N = 200$. Figure 1 compares the results obtained for $x = 0.7$ to the classical uncorrelated case ($x = 0$) [3] and to the solution provided by the saddle-point equations. We can notice that there is a good agreement between the simulations and the small x theory for small α . However, when $x \rightarrow 1$, the critical capacity computed from the saddle-points equations scales as

$$\alpha_c(x, \kappa = 0) \simeq \sqrt{\frac{1-x}{2\pi}} \exp\left(\frac{1}{2(1-x)}\right) \quad (22)$$

which is clearly unrealistic since the quantity of information $I(x)$ would diverge.

4.2 THE FERROMAGNETIC STRUCTURE OF THE SYNAPTIC WEIGHTS. — In order to shed some light on the way the network increases its storage capacity, we compute the first moments of the distribution of its weights (which is gaussian for small x) on the critical line $\alpha_c(\kappa)$

$$\overline{\langle J_{1j} \rangle} = A_c \sum_{k=2}^N C_{1k} \left[(\tilde{C} + \nu_c I)^{-1} \right]_{kj} \quad ; \quad A_c = \frac{\nu_c}{\kappa} \quad (23)$$

$$\overline{\langle J_{1j} J_{1k} \rangle} - \overline{\langle J_{1j} \rangle} \cdot \overline{\langle J_{1k} \rangle} = \frac{B_c}{N} \left[\frac{\tilde{C}}{(\tilde{C} + \nu_c I)^2} \right]_{jk} \quad ; \quad B_c = \frac{s_c \kappa + \nu_c (\kappa - m_c)}{\alpha_c \kappa H \left(\frac{m_c - \kappa}{\sqrt{s_c}} \right)} \quad (24)$$

In the usual uncorrelated case ($C_{ij} = \delta_{ij}$), all the synapses are of order $\frac{1}{\sqrt{N}}$, with zero mean. When the input data become spatially correlated, the weights J_{ij} are equal to the sum of two contributions :

- * a ferromagnetic term J_{ij}^+ : the mean values of the synapses linking two close cells (in the sense that $|i - j|$ is not much larger than the correlation length of the input data) are finite and positive. They decrease quickly with the distance separating the units but do not vanish when N goes to infinity.

- * a spin-glass term $J_{ij}^{s.g.}$. equation (24) shows that the couplings fluctuate around these mean values from sample to sample. Furthermore, the typical deviations are of order $\frac{1}{\sqrt{N}}$, like with uncorrelated patterns.

The origin of the first term J_{ij}^+ has already been explained in part 3.1. The parameter m_c may be interpreted as the typical value of the ferromagnetic part of the stabilities in the limit of small x .

For more general correlation matrices (i.e. not close to the identity), the replica calculations of part 3.1 are no longer valid. We can nevertheless predict the qualitative behaviour of the network. With respect to the usual uncorrelated case, the weights fluctuate around a ferromagnetic background J_{ij}^+ . The latter is self-averaging : it depends only on the statistical distribution C_{ij} of the patterns, and not on the particular training set. This positive synaptic "bump" provides all the units of the network with an additional field whose mean is equal to

$$h_+ = \sum_{j=2}^N C_{1j} J_{1j}^+ \quad (25)$$

for each pattern. As a result, the effective field due to the tuning of the weights (fluctuating parts) is $h_{s.g.} = \kappa - h_+$ rather than κ . Near the saturation (when $\kappa = 0$ for instance), $h_{s.g.}$ is negative, while the total field $h_{s.g.} + h_+$ is still positive. This explains how the critical capacity of the present model may exceed 2 [3], as found in part 3.2.

The results of simulations done with $C_{ij} = (0.7)^{|i-j|}$ and $\alpha = 1$ are displayed in figure 2.

The mean values of the weights $J_{ij}^+ = \overline{J_{ij}}$ and the typical fluctuations $J_{ij}^{s.g.} = \sqrt{J_{ij}^2 - \overline{J_{ij}}^2}$ are plotted for the distances $k = 1, \dots, 5$ and the different sizes $N = 50, 100$ and 200 (the numbers of samples the data are averaged over range from 10 to 100). The standard deviations seem to vanish like $\frac{B}{\sqrt{N}}$ as it is predicted in formula (24). Moreover they are roughly independent on the distance $|i - j|$ separating the units. Other simulations performed from $\alpha = 0.5$ to $\alpha = 2$ show that the coefficient B does not vary a lot when the size of the training set changes ($0.7 < B < 0.9$ typically). The mean couplings J^+ depend very weakly on the number of neurons N (the variations are taken into account by the sizes of the squares). However, J_{ij}^+ decreases quickly with $|i - j|$. Figure 3a shows that this decreasing is indeed exponential. It may be well fitted by the following law

$$J_{ij}^+ = J_0^+ y(x, \alpha)^{|i-j|} \quad (26)$$

Figure 3b displays the values of $y(x, \alpha = 0.5)$ obtained from figure 3a and compares them to the theoretical expectations. Equation (23) also predicts an exponential decreasing of the synaptic weights. We see that the agreement is quite good although the theory is exact up to order x^2 only. In addition to the statistical fluctuations, the error bars appearing in figure

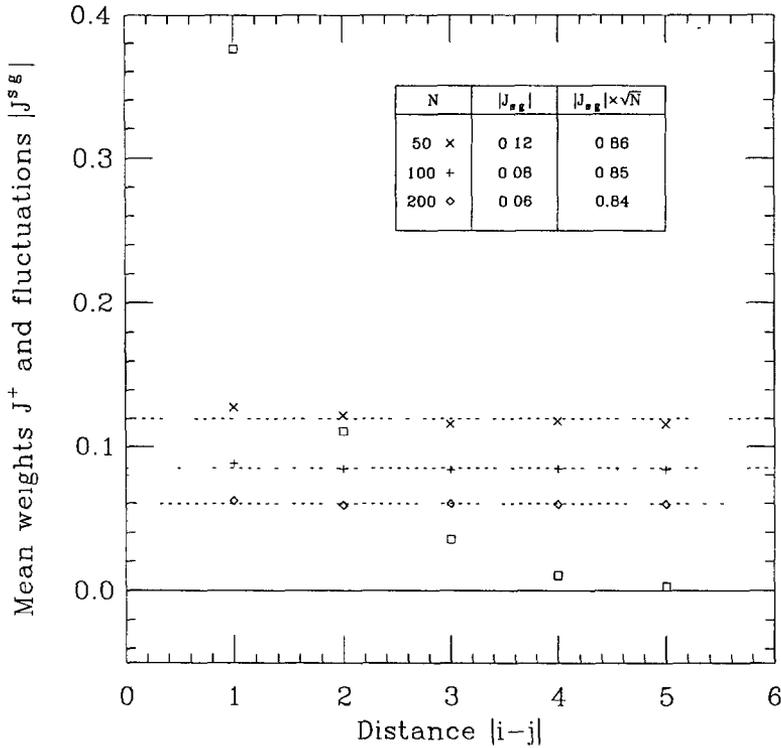


Fig.2. — the mean couplings $J_{ij}^+ = \overline{J_{ij}}$ (small squares) and the fluctuations $J_{ij}^{sg} = \sqrt{J_{ij}^2 - \overline{J_{ij}}^2}$ for different sizes $N = 50, 100$ and 200 and $\alpha = 1$. The distances $|i - j|$ range from 1 to 5. The scaling of the deviations is compatible with a $\frac{1}{\sqrt{N}}$ law. The mean weights seem to decrease exponentially with the distance separating two neurons (see figure 3). The size of the squares gives an upper bound of the statistical fluctuations of the J^+ (from sample to sample and inside each sample, from perceptron to perceptron) and of their variations for the different N .

3b include the uncertainty due the finite running time of the Minover. Using the notations of [7] and [4], one must take into account the performance guarantee factor A_T . It compares the largest stability κ_∞ achievable for one given sample (i.e. after an infinite time) to κ_T obtained by the algorithm at time T : $\kappa_T < \kappa_\infty < A_T \kappa_T$. All the simulations were performed with $1.02 < A_T < 1.06$.

Figure 4 displays the behaviour of J_{ij}^+ for a given $x = 0.7$ and different sizes of the training set. There is a good qualitative agreement with the Gaussian theory. For low α , the latter predicts

$$J_{ij}^+ \simeq \sqrt{\alpha} C_{ij} \quad (\alpha \rightarrow 0) \tag{27}$$

This may be easily understood, since the Hebb rule becomes optimal (for the stability criterion) for finite P . Thus, at low storage, the couplings reflect the inner structure of the input patterns. Near the saturation, the theoretical result is

$$J_{ij}^+ \simeq -\gamma (C^{-1})_{ij} \quad (\alpha \rightarrow \alpha_c) \tag{28}$$

where γ is a positive constant derived from the saddle-point equations. For $C_{ij} = x^{|i-j|}$,

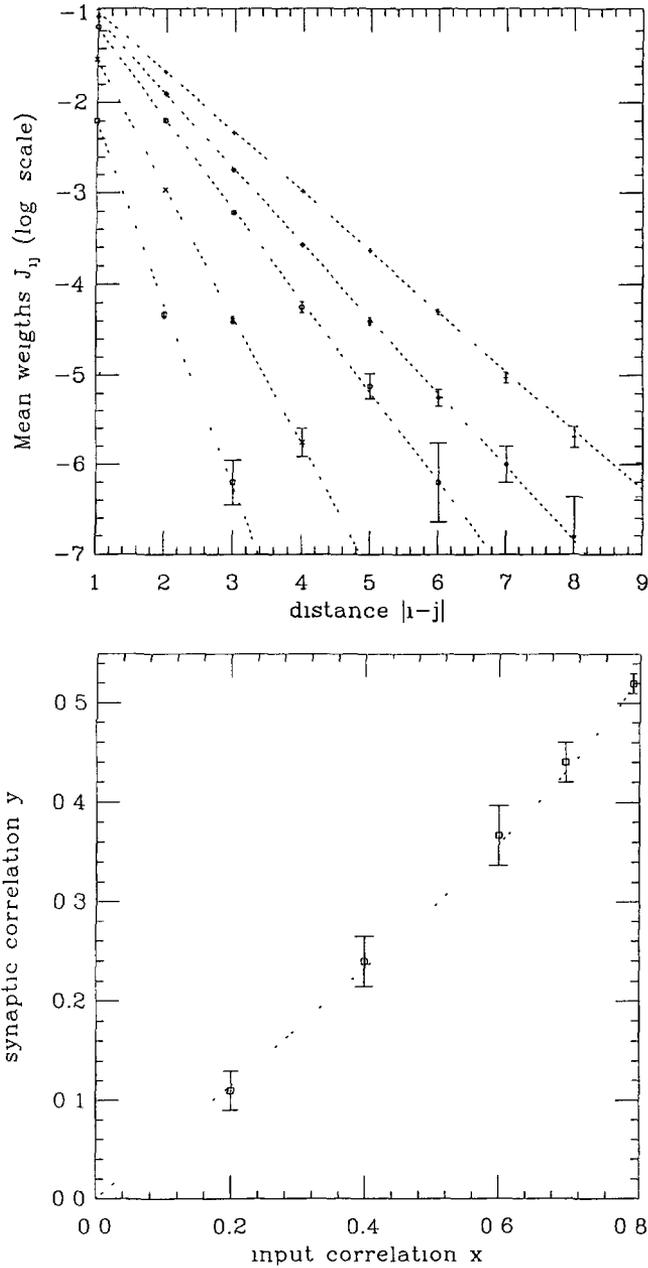


Fig.3. — (a) the mean values of the couplings J_{ij}^+ as a function of the distance $|i-j|$ (semi-log plot). The networks includes $N = 200$ neurons and the data are averaged over 100 samples. The size of the training set is $\alpha = 0.5$ and the different correlation strengths are from top to down : $x = 0.8, 0.7, 0.6, 0.4$ and 0.2 . The numerical data provided by the Minover algorithm seem to obey an exponential law proportional to $y^{|i-j|}$ (see equation (26)). (b) the dotted line shows $y(x)$ for $\alpha = 0.5$ as obtained from the weak correlation theory. The points are deduced from the figure 3a. The error bars take into account the statistical fluctuations and the uncertainty due to the stop of the Minover algorithm after a finite time (see part 4.2).

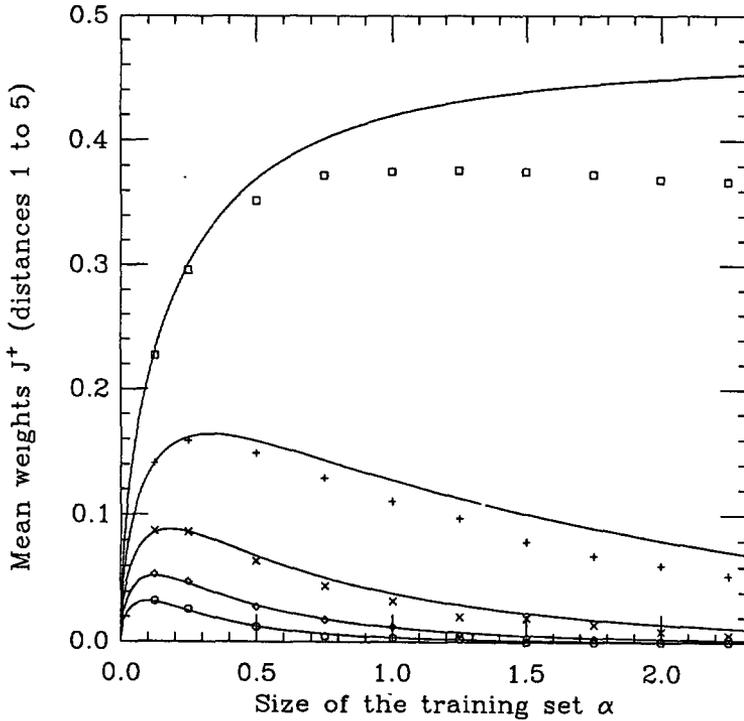


Fig.4. — the mean couplings J_{ij}^+ as functions of the training set size α for $x = 0.7$. From top to down, $|i - j| = 1, 2, 3, 4$ and 5 . The small x predictions (full curves) are compared to the simulation results (the size of the symbols represent the largest statistical standard deviation). Note the qualitative agreement for small α ($J_{ij}^+ \simeq \sqrt{\alpha} x^{|i-j|}$) and for large α ($J_{ij}^+ \ll 1$ if $|i - j| \neq 1$).

the only subsisting ferromagnetic couplings link the nearest neighbours [4]. This prediction (i.e. the mean weights are given by the inverse matrix at the critical capacity) seems to be corroborated by the simulations.

4.3 THE DYNAMICAL STABILITY OF THE PATTERNS. — The question we consider now is the following : does the ferromagnetic structure of the weights influence the retrieval abilities of the network ? Let us start from a stored pattern μ . We then choose a site at random (numbered l) and flip the corresponding spin. In the usual uncorrelated case, the stabilities of the other spins are changed by $O(\frac{1}{\sqrt{N}})$. In the large N limit, this single flip cannot affect the other units and the training pattern is perfectly retrieved through the dynamical evolution of the network (the notion of attraction basins is classically defined from the reversal of a finite fraction of the total number of neurons N [8]).

When the couplings have ferromagnetic means, the situation is more complicated. The stored pattern μ exhibit spatial correlations whose typical length is L with $C_{ij} = \exp(-|i - j|/L)$ (e.g. $L \simeq 2.8$ for $x = 0.7$ in the previous simulations). It may be roughly described as the juxtaposition of domains of same sign spins and the lengths of which are around L . Let us suppose now that the site $l + 1$ (or $l - 1$) belongs to the same domain as l . After the flip, its

stability is decreased by

$$\delta\Delta_{l+1}^\mu = -2J_{l,l+1}\xi_l^\mu\xi_{l+1}^\mu = -2J_{l,l+1} \quad (29)$$

which becomes equal to $-2J_{12}^+$ in the large N limit. As soon as α is large enough (i.e. $\kappa < 2J_{12}^+$), the neighbours are in turn flipped and the initial perturbation propagates. However, when it reaches the edge of the domain, the stability of the first neuron m of the opposite spins block is changed by

$$\delta\Delta_m^\mu = -2J_{m-1,m}\xi_{m-1}^\mu\xi_m^\mu = +2J_{m-1,m} \quad (30)$$

which is now positive. This neuron is thus stable. We see that the the initial spin flip cause the reversal of the whole domain this unit belonged to. We can conclude that the fixed point of the dynamics differs from the pattern μ in a finite number of bits (roughly L). There is no perfect retrieval but for large networks, the overlap between the two patterns goes to 1.

The above reasoning implicitly takes only into account the nearest neighbours interactions. It is justified by the sharp decreasing of the J^+ with the distance separating the units (see Fig.2). This discussion also pointed out the main difference between the spins belonging to centers of the domains (i.e. having the same sign as their neighbours) and the spins located at the edge of the blocks. The "center" units are, roughly speaking, stabilised by the J^+ couplings while the field incoming onto the "edge" spins due to the ferromagnetic weights vanishes. The latter can only be stored in their right directions (up or down) by the tuning of the J^{s-g} couplings. Since the number of such "edge" spins is a fraction of N (as soon as $L > 0$), the storage capacity increases.

5. Summary and discussion.

In this paper, we have focused on the storage properties of auto-associative memories when they are presented with spatially correlated patterns. We have computed the free energy for weak correlations. It already predicts two main differences with respect to the classical uncorrelated case.

First, the storage capacity increases but the total quantity of information seems to decrease. This situation is reminiscent of the biased patterns case [3] and emphasizes the necessity of decorrelating the data to improve the efficiency of the memories. One must however be cautious about the possible analogy with neurobiological facts (decorrelation of the visual stimuli through the first retina layers for instance). The results we obtained here depend strongly on the binary nature of the neurons. In particular, the ferromagnetic weights are useful since they allow the units to strengthen the input fields in the right direction. Such a principle would not be adequate any longer for continuous neurons (which may be more relevant for biological modelizations), where the intensity of the fields (and not only their signs) matters to the cells responses.

Secondly, the synaptic weights exhibit a richer structure than for uncorrelated patterns. In addition to the usual $O(\frac{1}{\sqrt{N}})$ fluctuating weights (i.e. tuned during the training), there is a $O(1)$ self-averaging and ferromagnetic background. These short range couplings take advantage of the spatial correlations of the input patterns to enhance the cells fields, without affecting too much the retrieval performances of large networks. Some numerical simulations were performed for exponentially decreasing correlations and one dimensional patterns. They show a surprisingly good agreement with above analytical predictions (valid for weak spatial correlations).

All the results we presented here were derived for continuous weights. It would be interesting to know what happens when other prescriptions are imposed over the couplings. With binary

weights ($J_{ij} = \pm \frac{1}{\sqrt{N}}$) or bounded synapses ($|J_{ij}| < \frac{J_0}{\sqrt{N}}$), the ferromagnetic means could not be of order 1 and the nature of the optimal synaptic matrix would be different.

Acknowledgments.

I am grateful to M. Mézard for numerous illuminating discussions and readings of the manuscript I also thank W. Krauth for useful remarks concerning the numerical simulations and I. Kocher, D. O'Kane, A. Treves for their help during the completion of this work.

References

- [1] J. J. Hopfield, *Proc Natl Acad Sci USA* **79** (1982) 2554.
- [2] D. Amit, "Modeling Brain Function" *Cambridge University Press* (1989).
- [3] E. Gardner, *J Phys A* **21** (1988) 257;
E. Gardner, B. Derrida, *J Phys A* **21** (1988) 271
- [4] R. Monasson, *J Phys A* **25** (1992) 3701.
- [5] J. J. Atick, A. N. Redlich, *Neural Comput.* **2** (1990) 308.
- [6] J. P. Nadal, N. Parga, "Information processing by a perceptron" *Proceedings of ELBA, Neural Networks from Biology to High Energy Physics* (1992).
- [7] W. Krauth, M. Mézard, *J Phys A* **20** (1987) L745.
- [8] T. B. Kepler, L. F. Abbott, *J Physique* **49** (1988) 1657;
B. M. Forrest, *J Phys A* **21** (1988) 245.
- [9] D. Amit, M. Evans, H. Horner, K. Y. M. Wong, *J Phys A* **23** (1990) 3361;
M. Griniasty, H. Gutfreund, *J Phys A* **24** (1991) 715.