# Dynamical Learning Process for Recognition of Correlated Patterns in Symmetric Spin Glass Models

EUROPHYSICS LETTERS 1 November 1987

*Europhys. Lett.*, 4 (9), pp. 979-985 (1987)

# Dynamical Learning Process for Recognition of Correlated Patterns in Symmetric Spin Glass Models.

G. PÖPPEL and U. KREY

*Institut für Physik III der Universität Regensburg - D-84 Regensburg, F.R.G.*

(received 13 April 1987; accepted in final form 17 August 1987)

PACS. 05.90. – Other topics in statistical physics and thermodynamics.

**Abstract.** – In the framework of spin glass models with symmetric interactions a local dynamical learning process is studied, by which the energy landscape is modified in such a way that even strongly correlated noisy patterns can be recognized. Additionally the basins of attraction of the patterns can be systematically enlarged. After completion of the learning process the system can recognize as many patterns as there are neurons ($p = N$), and for small systems even more ($p > N$). The dependence of the learning time $R$ on the parameters of the system (*e.g.*, the average correlation, the noise level, and the number $p$ of patterns) is studied and it is found that $R$ increases as $p^x$, with $x \approx 3.5$, as long as $p < N$, whereas for $p > N$ the increase is more drastic.

## 1. Introduction.

Pattern recognition of spin glass models is by now already a rather well-established field [1]; however, the standard Hebb model (see below) has certain weaknesses, since i) only orthogonal or uncorrelated random patterns can be recognized, and ii) the ratio $\alpha$ between the number $p$ of patterns and the number $N$ of neurons must be smaller than $\simeq 0.14$ [2, 3].

In the present letter we show that these shortcomings can be remedied by an iterative error-correcting learning process, through which the Hebb interaction matrix is modified. Our learning process is dynamic, *i.e.* in contrast to an established «static» prescription [4, 5], by which a suitable interaction matrix for the recognition of correlated patterns is calculated algebraically and in a nonlocal way from the totality of all patterns, our learning process is based on the repeated application of a natural and rather simple local error correction scheme involving one pattern at a time. Notably, our interaction matrix is always symmetric, both in the course of the learning process and after its completion. This means that the energy is always well defined and that the relaxation process (which is based on the usual single spin flip Monte Carlo dynamics [6]) leads at $T = 0$ to a nearby local minimum or to a stationary thermodynamic equilibrium at $T \neq 0$.

In fact, the basic ingredient of our learning process is a natural modelling of the energy landscape (see below), through which the pattern states become local minima of the system, whereas unwanted minima (*i.e.* «spurious patterns») are energetically enhanced. Of course, within the realization of this concept there is a certain amount of arbitraryness, *e.g.*,

concerning the depth of the minima and the height of the barriers between them. However, what is perhaps more important, even the basins of attraction of the different patterns can be systematically enlarged by our learning process, simply by teaching the system to recognize the original patterns even «in disguise», *e.g.*, if they are perturbed by a finite amount of randomness («noisy patterns»). It should be stressed that in our model neither the correlations of the patterns, nor the amount of noise nor the above-mentioned ratio $\alpha = p/N$ must be small.

## 2. The learning algorithm.

We consider the usual Ising spin glass Hamiltonian

$$H = -\frac{1}{2} \sum_{j,k} J_{jk} S_j S_k \,, \tag{1}$$

where the indices $j, k = 1, \ldots, N$ denumerate the $N$-neurons. The two states of these neurons («firing» or «not firing») are represented by the Ising variables $S_j = \pm 1$, and the coupling constants $J_{j,k}$, which vanish for $j = k$, describe the mutual interactions of the neurons through synaptic links. Then there are $p$ different «patterns» $\xi^\mu$ out of the $2^N$ possible spin configurations; these patterns may be correlated, *i.e.* the overlap function $q_{\mu\nu} := (1/N)(\xi^\mu, \xi^\nu)$, where $(\xi^\mu, \xi^\nu)$ denotes the usual scalar product of real $N$-component vectors, can be different from zero. After the learning process, by which the $J_{j,k}$ are changed (see below), these patterns should be recognized by the system (1) through the usual sequential relaxation process [1].

To be specific, our relaxation procedure proceeds as follows: of three consecutive cycles, during the first and third cycle $N$-times a position $j$ is selected randomly and the spin $S_j$ is flipped if this leads to a lower energy, while during the second cycle the spins are visited sequentially. Thus we try to avoid that, on the one hand, some spins are incidentally overlooked, as might be the case by a completely random selection of the spins, while, on the other hand, we avoid any systematic bias, which might be produced, *e.g.*, in the learning process if the spins were always visited in the same order.

The learning process proceeds as follows: we start with the Hebb-Hamiltonian, *i.e.* with

$$J_{j,k} = \frac{1}{N} \sum_{\nu=1}^{p} \xi_j^\nu \xi_k^\nu \,. \tag{2}$$

Then, a random permutation $\nu_1, \ldots, \nu_p$ of the $p$ patterns is selected; starting with $\nu_1$, certain *input vectors* $\eta^{\nu_i}$ are submitted, one after the other, *i.e.* for $i = 1$ to $i = p$, to the relaxation process described above. These input vectors can be taken either as the original patterns or as some «noisy» modifications of them. For every $i$, as many relaxation cycles are performed as are necessary to get the system definitely trapped in a local minimum. This minimum state is the *output vector* $x^{\nu_i}$. If it is different from the original pattern $\xi^{\nu_i}$, the Hamiltonian (1) is modified as follows: $J_{j,k} \to J_{j,k} + \Delta J_{j,k}$, for all pairs $(j, k)$, with

$$\Delta J_{jk} = -\frac{\lambda}{N} (x_j^{\nu_i} x_k^{\nu_i} - \xi_j^{\nu_i} \xi_k^{\nu_i}) \,, \tag{3}$$

and the relaxation of the next input vector proceeds with the new Hamiltonian. In (3), $\lambda$ is a positive number determining the strength of the correction, and implicitly the speed of the learning process (see below).

The teaching process is stopped, if for all members of the permutation considered the output vectors $\boldsymbol{x}^\nu$ are identical with the patterns $\boldsymbol{\xi}^\nu$ to be learned; otherwise it is repeated with a new permutation and with new noisy modifications of the corresponding patterns. Of course, one always uses the latest version of the corrected Hamiltonian.

In general, a very large number of iterations is necessary, particularly if $p$ is comparable to $N$ (see below), however, in practice we have always found that after a sufficient number of iterations the learning process stopped, although we do not yet have a proof for this fact. (For a different learning process with an *asymmetric* $J_{j,k}$ matrix, a convergence proof has been given in [7].)

In any case, one can easily interprete the physics behind our procedure by considering the energy change:

$$\Delta H(\boldsymbol{S}) = -\frac{1}{2}\sum_{j,k} S_j \Delta J_{jk} S_k = \frac{\lambda}{2N}[(\boldsymbol{x}^{\nu_i}, \boldsymbol{S})^2 - (\boldsymbol{\xi}^{\nu_i}, \boldsymbol{S})^2] , \tag{4}$$

induced by the correction (3) for a given spin configuration $\boldsymbol{S}$.

For $\boldsymbol{S} = \boldsymbol{x}^{\nu_i}$, *i.e.* for the «unwanted» output resulting from the input vector corresponding to pattern $\boldsymbol{\xi}^{\nu_i}$ (*e.g.*, a noisy modification), $\Delta H$ is positive, namely

$$\Delta H(\boldsymbol{x}^{\nu_i}) = \frac{\lambda N}{2}[1 - q^2(\boldsymbol{x}^{\nu_i}, \boldsymbol{\xi}^{\nu_i})] , \tag{5}$$

while for $\boldsymbol{S} = \boldsymbol{\xi}^{\nu_i}$, $\Delta H$ is negative, namely $\Delta H(\boldsymbol{\xi}^{\nu_i}) = -\Delta H(\boldsymbol{x}^{\nu_i})$.

Thus, the essential point of our learning process consists in a systematic increase (decrease) of the energy of unwanted (wanted) states, *i.e.* the energy landscape in the vicinity of the patterns $\boldsymbol{\xi}^\nu$ is modelled in such a way that these patterns become local minima. Moreover, by performing the learning process not with the original patterns, but with noisy modifications of them, one can systematically enlarge the basins of attraction of the patterns. At the same time, the probability to have unwanted (*i.e.* «spurious») minima should be strongly reduced.

## 3. Results.

In fig. 1 results are presented, which characterize the improvement of the recognition process through the learning procedure for a system with $N = 100$ neurons and $p = 10$ patterns with an averaged correlation of $\overline{q_{\mu\nu}} = 0.118 \pm 0.087$: the *retrieval quality*, *i.e.* the overlap $q(\boldsymbol{\xi}^1, \boldsymbol{x}^1)$ of the original pattern $\boldsymbol{\xi}^1$ and the stationary output $\boldsymbol{x}^1$ of the relaxation is plotted over the noise level $p_n$ (fraction of randomly flipped spins). Every point on the curves of fig. 1 represents an average over 100 different noisy modifications of pattern $\boldsymbol{\xi}^1$ with identical noise level, and the error bars represent the standard deviation from the average. (Actually the distribution of the 100 results for a given by $p_n$ is strongly non-Gaussian, with a sharp peak at the maximum value of $q(\boldsymbol{\xi}^1, \boldsymbol{x}^1)$ in each curve, *i.e.* at 0.55 in fig. 1*a*), and at 1 in fig. 1*b*), and a broader distribution centred around a smaller value, appearing for noise levels $\geq 0.3$.) In any case, from a comparison of fig. 1*a*) with fig. 1*b*) it is obvious that the retrieval quality has been drastically improved through the learning process (fig. 1*b*)), *i.e.* after a learning process which has been performed with *noisy* patterns, with a noise level $\sigma = 0.3$.

If the learning process would be performed with pure patterns ($\sigma = 0$), the retrieval quality would still be much better than in fig. 1*a*), but not as good as in fig. 1*b*). The reason is

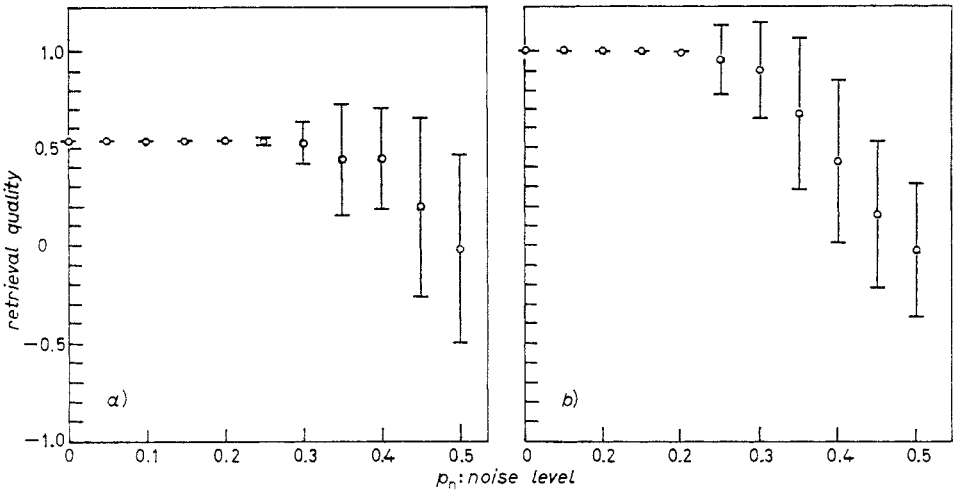


Fig. 1. – The retrieval quality, *i.e.* the averaged overlap $q(\xi^1, \boldsymbol{x}^1)$ of the original pattern $\xi^1$ and the output $\boldsymbol{x}^1$ of the relaxation process, is presented for 100 relaxation processes starting with different noisy modifications of the original pattern, as a function of the noise level $p_n$, *i.e.* the relative number of randomly flipped spins of the modifications, for a system with $N = 100$ neurons and $p = 10$ patterns, which have an averaged correlation of $\overline{q_{\mu\nu}} = 0.118 \pm 0.087$. Figure 1*a*) is for the Hebb system, eq. (2), *i.e.* without learning, fig. 1*b*) for the system as obtained after a learning process where additionally the basins of attraction have been enlarged by using noisy input patterns with a noise level of $\sigma = 0.35$. The learning strength was $\lambda = 0.0005$, and the total number $R$ of learning steps was 5222 in case *b*).

that for $\alpha = 0.3$, in contrast to $\sigma = 0$, there is an additional enlargement of the basins of attraction by the learning process.

Furthermore, using a sufficiently long teaching time, we have found that for small systems 100% retrieval can be obtained even for *more* than $N$ patterns, *i.e.* for $p = 60$ uncorrelated patterns in case of $N = 50$, see below.

Of course, the necessary number $R$ of *teaching steps* (*i.e.* corrections of the Hamiltonian) increases strongly with $p$, $\alpha$, $\overline{q_{\mu\nu}}$ and $1/\lambda$.

We have found that $R$ is proportional to $1/\lambda$ and proportional to $\overline{q_{\mu\nu}}$ at least for $0.01 \leqslant \lambda \leqslant 0.08$ and $0.10 \leqslant \overline{q_{\mu\nu}} \leqslant 0.20$, with $R \simeq 400$ for $\overline{q_{\mu\nu}} = 0.1$ and $\lambda = 0.01$ (with $N = 100$, $p = 20$).

The dependence on the number of patterns can be seen from fig. 2*a*). For systems with $N = 50$, 100 or 200 neurons the number $R$ of learning steps for correct recognition of $p$ uncorrelated patterns increases strongly with $p$: however, for $p < N$ the increase seems to be nonexponential, *e.g.*, for $N/2 < p < N$, $R$ is found to behave as $\sim p^x$, with $x \approx 3.2$ for $N = 100$ and $\approx 3.6$ for $N = 50$. Only for $p \geqslant N$, $R$ increases more drastically; however, from our data we cannot make a definite statement whether, *e.g.*, for $N = 50$ and $50 \leqslant p \leqslant 60$ the increase is $\sim p^y$ with $y \approx 9$, or whether the increase is even exponentially. In fact, for a generic spin glass, *i.e.* with a Gaussian exchange, one has *exponentially* many local minima [8] and would of course expect an exponentially large teaching time. Thus one may speculate that the models prepared by our teaching process may somehow interpolate between the separable Hebb model and the generic case, and that the cross-over happens around $\alpha \simeq 1$.

Finally, in fig. 2*b*), the dependence of the learning time on the averaged correlation $\overline{q_{\mu\nu}}$ is

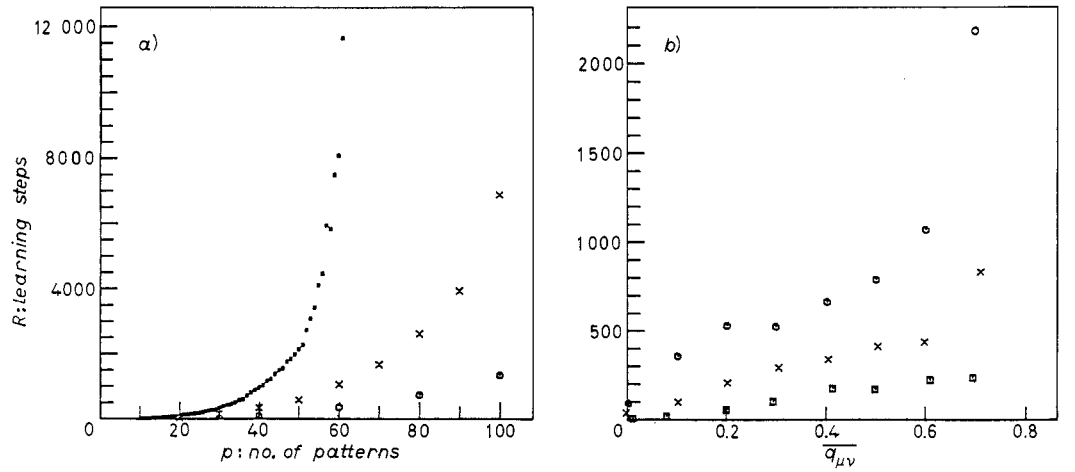


Fig. 2. – The number of learning steps $R$ is studied as a function of various system parameters, namely in fig. 2*a*) as a function of $p$ up to very large values of $p$, for $\lambda = 0.04$ and $\overline{q_{\mu\nu}} = 0$ (circles: $N = 200$, crosses: $N = 100$, squares: $N = 50$); in fig. 2*b*) as a function of $\overline{q_{\mu\nu}}$ up to very large correlations, again with $\lambda = 0.04$ (circles: $N = 200$, $p = 40$; crosses: $N = 100$, $p = 20$; squares: $N = 50$, $p = 10$).

studied. Again, the increase is not very drastic, except around extremely large values, *i.e.* for $\overline{q_{\mu\nu}} > 0.6$.

As already mentioned, for the enlargement of the basins of attraction it is necessary to perform the learning process with noisy modifications of the original patterns. In fig. 3 we study the question, whether the corresponding noise level $\sigma$ prolongs the learning time. As can be seen from fig. 3, for $\sigma < 0.2$ this is practically *not* the case, and also for the very large value $\sigma = 0.3$, the increase of $R$ compared with $\sigma < 0.2$ amounts only to a factor 2.

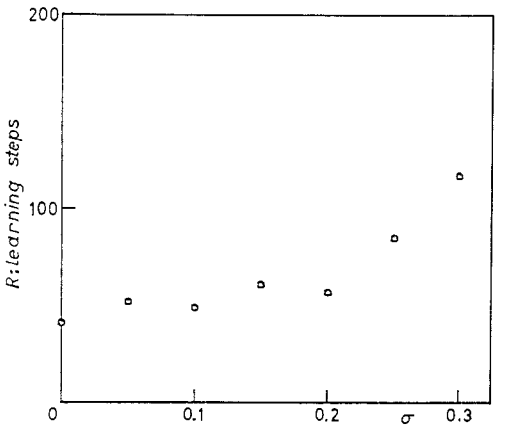


Fig. 3. – The total number of learning steps $R$ for the learning process with *noisy* patterns (cf. fig. 1*b*)) is presented over the noise level $\sigma$. $N = 100$, $p = 10$, $\lambda = 0.02$, $\overline{q_{\mu\nu}} = 0.118 \pm 0.087$.

To make the retrievel quality and the possible applications obvious, we present in the final fig. 4 a recognition process, where in a system with $N = 256$ neurons $p = 6$ extremely correlated patterns, namely the letters $A$, $B$, $C$, $D$, $E$, $F$, which have an averaged correlation of 0.78, are recognized after just three relaxation cycles, although these patterns

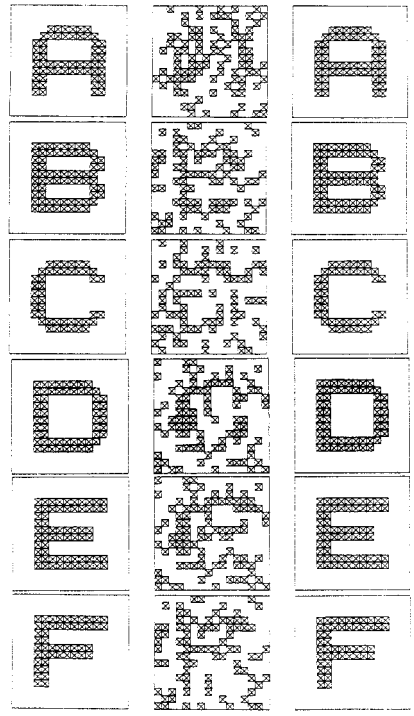

Fig. 4. – Pattern recognition of the letters $A$, $B$, $C$, $D$, $E$, $F$, as explained in the text. The first column represents the pure patterns, the second column the noisy modifications, which are taken as the input vectors of the recognition process, and the third column the output of the recognition, which took three relaxation cycles. (Learning parameters: $\lambda = 0.05$, $\sigma = 0.3$, $R = 333$.)

are presented to the system «in strong disguise» (see the second column), corresponding to a noise level of $p_n = 0.3$, so that the human eye would no longer recognize them. The number of learning steps, which were performed with $\lambda = 0.05$ and $\sigma = 0.3$, was $R = 333$. Of course, still stronger «disguise» of the patterns would hardly make sense in the present example. In any case at this place we would like to stress that it is not primarily the Hamming distance, but rather the sculpturing of the energy landscape, *i.e.* our learning process, which determines whether, *e.g.*, the first noisy pattern is recognized as an «$A$» and not as a «$B$».

## 4. Conclusions.

We have introduced an energy-preserving learning process, by which a system of $N$ mutually interconnected neurons can be prepared in such a way that up to $\simeq N$ strongly correlated and rather noisy patterns can be successfully recognized in extremely short time. The dependence of the learning time on various parameters has been studied in detail and it has been found that reasonable results can be obtained with reasonable (*i.e.* non-exponentially large) effort. The concept of our learning process can of course also be applied to relaxation at finite $T$, to different recognition criteria, and with different Hamiltonians.

After the completion of the present work, we received a preprint [9] by which we learned that similar learning algorithms for symmetrical couplings $J_{lm}$, although with a correction prescription differing from our eq. (3), are being studied within the Edinburgh group [9, 10].

* * *

The authors would like to thank S. DIEDERICH, D. W. HEERMANN, H. HORNER and J. P. NADAL for hints and stimulating discussions.

## REFERENCES

[1] See, *e.g.*, the lectures of AMIT D., KINZEL W., TOULOUSE G., DREYFUS G., VAN HEMMEN L., HERTZ J. and SOMPOLINSKY C., *Conference on Glassy Dynamics and Optimization, Heidelberg 1986*, to appear in *Lecture Notes in Physics* (Springer Verlag, Berlin-Heidelberg-New York) 1987.
[2] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. A*, **32** (1985) 1007.
[3] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. Lett.*, **55** (1985) 1483.
[4] PERSONNAZ L., GUYON I. and DREYFUS G., *J. Phys. (Paris) Lett.*, **46** (1985) L-623.
[5] PERSONNAZ L., GUYON I. and DREYFUS G., *Phys. Rev. A*, **34** (1986) 4287.
[6] BINDER K. (Editor), *Monte Carlo Methods*, in *Statistical Physics*, 2nd Edition (Springer Verlag, Berlin-Heidelberg-New York) 1986.
[7] DIEDERICH S. and OPPER M., *Phys. Rev. Lett.*, **58** (1987) 949.
[8] BRAY A. and MOORE M. A., *J. Phys. C*, **13** (1980) L-469.
[9] BRUCE A. D., CANNING A., FORREST B., GARDNER E. and WALLACE D. J., *Neural Networks for Computing*, edited by J. S. DENKER (Snowbird, Utah). AIP *Conference Proceeding*, Vol. **151**, (American Institut of Physics) 1987.
[10] GARDNER E., STROUD N. and WALLACE D. J., Edinburgh preprint 87/394 (submitted to *Phys. Rev. Lett.*).