Learning of Correlated Patterns in Spin-Glass Networks by Local Learning Rules

Sigurd Diederich and Manfred Opper
Institut für Theoretische Physik, Universität Giessen, D-6300 Giessen, Federal Republic of Germany
(Received 18 November 1986)

Two simple storing prescriptions are presented for neural network models of N two-state neurons. These rules are local and allow the embedding of correlated patterns without errors in a network of spin-glass type. Starting from an arbitrary configuration of synaptic bonds, up to N patterns can be stored by successive modification of the synaptic efficacies. Proofs for the convergence are given. Extensions of these rules are possible.

PACS numbers: 87.30.Gy, 06.50.Mk, 75.10.Hk, 89.70.+c

In recent years models of neural networks have become increasingly popular.[1-6] Highly organized performance characteristics of the brain, such as fault-tolerant content-addressable memory, appear as a spontaneous collective property of a system of interconnected neurons. Implementing algorithms for these networks, which allow the utilization of these associative properties for problem solving, is a challenging task in the context of artificial intelligence.

The neural networks are typically modeled as a system of N totally interconnected neurons, each having the states $S_i=1$ (firing) or $S_i=-1$ (quiescent). Neuron i, the postsynaptic neuron, receives the potential $J_{ij}S_j$ from neuron j, the presynaptic one, where J_{ij} characterizes the synaptic efficacy. The local field of neuron i is the sum of the contributions from all presynaptic neurons,

$$E_i=\sum_{j=1}^{N} J_{ij}S_j. \tag{1}$$

The dynamics of the system is defined by the deterministic asynchronous spin-flip algorithm used in zero-temperature Monte Carlo simulations. Stationary states $\mathbf{S}=(S_1,\ldots,S_N)$ of the dynamics are characterized by

$$S_iE_i=\sum_{j=1}^{N} S_iJ_{ij}S_j>0,\quad i=1,\ldots,N. \tag{2}$$

For the network to serve as a content-addressable memory, the set of prescribed patterns to be memorized has to be embedded by the application of some learning rule. This embedding procedure leads to synaptic connectivities such that the patterns become local or approximate local attractors with respect to the dynamics of the system.

In biological as well as device contexts local learning rules are of importance.[7] A learning rule is local if the change of the synaptic efficiency J_{ij} depends only on the states of the interconnected neurons i,j and possibly on the postsynaptic local field E_i.[8] This corresponds to every step of the learning procedure. Hebb's rule and its modifications have become the most popular representatives of these local rules,[1,9-11] being able to provide storage of uncorrelated patterns without error. Although much effort has been devoted to improvement of their storing capabilities,[12-14] networks with local learning rules fail, up to now, in memorizing correlated patterns.[4,15,16] This point severely restricts their applicability. On the other hand, for arbitrary patterns it has been shown that the required synaptic couplings J_{ij} can be calculated by means of matrix inversions.[5,17] However, up to the present it has not been possible to substitute a local learning rule for this procedure.

The purpose of this Letter is to present two learning rules which are both local and able to yield storage of correlated as well as uncorrelated patterns without errors. Each pattern is learned by repeated presentation of it to the network in a sequence of learning steps. The learning process is terminated when all patterns $\mathbf{S}^\nu=(S_1^\nu,\ldots,S_N^\nu)$ fulfill the embedding condition,

$$S_i^\nu E_i^\nu \geq T,\quad i=1,\ldots,N, \tag{3}$$

with a positive threshold $T\sim1$, which has been introduced to ensure the local stability of a pattern $\mathbf{S}^\nu$. This local stability can easily be proven from Eq. (3) for a large network with $J_{ij}=O(1/\sqrt{N})$.[18] For simplicity we choose $T=1$ in the following. The prescribed patterns can be embedded into a network with arbitrary initial configuration of synapses, $J_{ij}=J_{ij}^0=O(1/\sqrt{N})$, e.g., a Sherrington-Kirkpatrick spin-glass.[19] Learning rule I which is of Perceptron type[20,21] leads the network to satisfy the set of inequalities (3) for all patterns $\mathbf{S}^\nu$ after a finite number of learning steps, whereas for rule II the magnitudes of all fields on the left-hand side of Eq. (3) converge asymptotically to the threshold value $T=1$. We give explicit proofs for the convergence of both rules under very weak conditions.

We first explain rule I: The learning process is started with pattern $\nu=1$ by checking whether this pattern is already embedded in the network, i.e., if the system (3) is obeyed for $\nu=1$. If for example (3) is not satisfied for neuron i, we update the synaptic bonds J_{ij} according to the rule $J_{ij} \to J_{ij} + \delta J_{ij}$ with

$$\delta J_{ij} = (N-1)^{-1} S_i^\nu S_j^\nu, \quad \text{for } j \neq i. \tag{4}$$

Self-couplings are always excluded, i.e., $J_{ii} \equiv 0$, throughout the learning process. If the condition (3) is satisfied for neuron i, we leave the bonds J_{ij} unchanged for $j=1,\ldots,N$. To proceed, we pass to pattern $\nu=2$ and update the bonds in the same way and so on up to $\nu=p$. This leads to a sequence of modifications of the synaptic efficacies which all together constitute one cycle of our learning procedure. We repeat these cycles again and again until the embedding condition is simultaneously satisfied for every pattern ν at every site i. Our procedure is in contrast to previous learning rules where each pattern was memorized in a single learning event. This seems to us to be the main reason why they were not able to give storage of correlated patterns. Clearly, not every presentation of a pattern ν results in a modification of the bonds. Instead, a learning step takes place only if the field $S_i^\nu E_i^\nu$ does not exceed the threshold. This is summarized in the equation

$$\delta J_{ij} = (N-1)^{-1} S_i^\nu S_j^\nu \theta(1 - E_i^\nu S_i^\nu), \quad j \neq i. \tag{5}$$

$\theta(x)$ is the Heaviside function. We shall prove in the following that after a finite number of steps the fields of all patterns will exceed the threshold value $T=1$.[22] Thus the learning process terminates.

The proof of the convergence closely follows arguments worked out more than two decades ago for the Perceptron.[23] Remarkably, this proof requires only the existence of a network for which the prescribed patterns $\nu=1,\ldots,p$ are stationary states, i.e., there should exist a matrix $\{J_{ij}^*\}$ with $J_{ii}^*=0$ and

$$\sum_{j=1}^{N} S_i^\nu J_{ij}^* S_j^\nu > 0, \quad i=1,\ldots,N, \quad \nu=1,\ldots,p. \tag{6}$$

Obviously, the learning rule (5) does not couple different rows i of the synaptic matrix. Therefore it is sufficient to work out the proof for a single row i. For simplicity of presentation we restrict ourselves to the case of a *tabula rasa* learning rule, i.e., $J_{ij}^0=0$, an empty network at the beginning. We introduce the following abbreviations:

$$I_j = J_{ij}, \quad \sigma_j^\nu = S_i^\nu S_j^\nu. \tag{7}$$

If x_β denotes the actual number of times for which pattern β has led to a modification of the synaptic bonds, i.e., to a learning step, we can write

$$I_j = (N-1)^{-1} \textstyle\sum_\beta x^\beta \sigma_j^\beta, \quad I_i = 0. \tag{8}$$

Equation (6) is written as

$$\sum_{j\neq i} I_j^* \sigma_j^\nu > 0. \tag{9}$$

Now we assume that some pattern α does not satisfy the embedding condition, i.e., a learning step has to be performed. We calculate the change of $\sum_j I_j^2$ resulting from this step. This yields

$$\Delta\left(\sum_{j\neq i} I_j^2\right) = (N-1)^{-1}\left(1 + 2\sum_{j\neq i} I_j \sigma_j^\alpha\right) < 3/(N-1) \equiv D.$$

Here we have used $\sum_{j\neq i} I_j \sigma_j^\alpha < 1$, because the embedding condition is not satisfied for pattern α. Since $\sum_{\beta=1}^{p} x^\beta$ is the total number of learning steps these changes add up to

$$\sum_{j\neq i} I_j^2 = \sum \Delta\left(\sum_{j\neq 1} I_j^2\right) < D \sum_{\beta=1}^{p} x^\beta$$

for the total process.

If we introduce $y^\beta \equiv x^\beta (\sum_{\nu=1}^{p} x^\nu x^\nu)^{-1/2}$ it follows that

$$\sum_{j\neq 1}\left[(N-1)^{-1}\textstyle\sum_\beta y^\beta \sigma_j^\beta\right]^2 = \displaystyle\sum_{j\neq 1} I_j^2 \left[\textstyle\sum_\nu x^\nu x^\nu\right]^{-1} < D \textstyle\sum_\beta x^\beta \left[\sum_\nu x^\nu x^\nu\right]^{-1} \leq Dp\left[\sum_\beta x^\beta\right]^{-1}. \tag{10}$$

The last bound is obtained from the Schwartz inequality. If we now assume that the learning process never terminates, i.e., $\sum_\beta x^\beta$ grows infinitely, we conclude that the left-hand side of (10) becomes arbitrarily small for $y^\beta, \beta=1,\ldots,p$ on the positive octant of the unit sphere. So by the compactness of this set the equations[24]

$$\textstyle\sum_\beta \xi^\beta \sigma_j^\beta = 0 \text{ for all } j \neq i$$

are fulfilled for some $\xi^\beta \geq 0$, $\beta=1,\ldots,p$, with $\sum_\beta \xi^\beta \xi^\beta = 1$. Multiplying each equation by I_j^* and summing over j we obtain

$$0 = \sum_\beta \xi^\beta \sum_{j\neq i} I_j^* \sigma_j^\beta.$$

This contradicts Eq. (9), since at least one ξ^β is positive. So in fact $\sum_\beta x^\beta$ is bounded and the learning process terminates.

The purpose of learning rule II is to modify the synaptic bonds in such a way that the fields of all embedded patterns become equal to 1:

$$E_i^\nu S_i^\nu = 1, \quad i=1,\ldots,N. \tag{11}$$

This property of the fields may serve to discriminate between memorized and spurious states of the system[25] in the pattern-recognition process. As in rule I all patterns are presented sequentially to the network. The synaptic coefficients are updated according to

$$\delta J_{ij} = N^{-1}(1 - E_i^\nu S_i^\nu) S_i^\nu S_j^\nu. \tag{12}$$

In contrast to Eq. (5), every presentation of a pattern leads to a synaptic modification. From (12) we see that if $\delta J_{ij} \to 0$, the synaptic coefficients satisfy the embedding condition (11). However, this equality is obtained only in the limit of an infinite number of learning cycles. We show in the following that indeed the procedure converges. For convenience we include self-interactions J_{ii}. In this case the convergence can be proven under the weak condition that the patterns are linearly independent. Thus up to N patterns can be stored. In case of $J_{ii}=0$ the proof is almost the same, but the sufficient conditions cannot be formulated as simply. We again restrict ourselves to one row of the synaptic matrix, start with a *tabula rasa* network, and use the abbreviations (7). Equation (12) is written

$$I_j \to I_j + N^{-1}\left[1-\sum_{k=1}^{N} I_k \sigma_k^\alpha\right]\sigma_j^\alpha.$$

Analogous to Eq. (8) (but change $N-1$ to N), we introduce the embedding strengths x^α which are no longer integers in this case. In the $(l+1)$th learning cycle these quantities change according to

$$\begin{aligned} x^\alpha(l+1)-x^\alpha(l) &= 1-\textstyle\sum_k I_k\sigma_k^\alpha \\ &= 1-N^{-1}\sum_{k,\beta} x^\beta \sigma_k^\beta \sigma_k^\alpha. \end{aligned} \qquad (13)$$

On the right-hand side of this equation all strengths x^β for $\beta<\alpha$ have been already updated, i.e., $x^\beta = x^\beta(l+1)$ for $\beta<\alpha$, whereas for $\beta \geq \alpha$ the x^β's take their previous values, i.e., $x^\beta = x^\beta(l)$ for $\beta \geq \alpha$. So

$$x^\alpha(l+1) = 1-\sum_{\beta<\alpha} B^{\alpha\beta}x^\beta(l+1)-\sum_{\beta>\alpha} B^{\alpha\beta}x^\beta(l), \qquad (14)$$

with the symmetric matrix $B^{\alpha\beta}=N^{-1}\sum_k \sigma_k^\alpha \sigma_k^\beta$. Assuming, now, that the limit $x^\beta(l\to\infty)$ exists for $\beta = 1,\ldots,p$, we immediately see that $x^\beta(\infty)$ obeys the system of linear equations

$$\sum_{\beta=1}^{p} B^{\alpha\beta}x^\beta(\infty)=1, \quad \alpha=1,\ldots,p. \qquad (15)$$

Equation (14) is the Gauss-Seidel iterative method[26] for solving the linear system (15). This method converges if the symmetric matrix B is positive definite, i.e., the expression

$$\sum_{\alpha,\beta} y^\alpha B^{\alpha\beta} y^\beta = \frac{1}{N}\sum_j \Big(\sum_\alpha y^\alpha \sigma_j^\alpha\Big)^2$$

is strictly positive for every configuration y^α ($\alpha=1,\ldots,p$) which does not vanish identically. This condition is always fulfilled, because the patterns $\mathbf{S}^\alpha$ are assumed to be linearly independent. The limiting values ($l\to\infty$) of the synaptic coefficients can be calculated analytically. Solving Eq. (15) and restoring the original denotations, we find

$$J_{ij} = N^{-1}\sum_{\beta,\gamma}(C^{-1})^{\beta\gamma}S_i^\beta S_j^\gamma, \qquad (16)$$

with

$$C^{\alpha\beta}=N^{-1}\sum_{j=1}^{N} S_j^\alpha S_j^\beta.$$

The synaptic connectivities (16) are identical to those obtained earlier by nonlocal rules.[5]

We have demonstrated in this Letter how to overcome the limitations connected with the storing of correlated patterns with local learning rules. Two simple storing prescriptions have been presented which allow for rigorous proofs of their convergence. Obviously, a whole class of more sophisticated rules may be obtained by generalizations. From a biological point of view, for example, one may prefer a multiplicative procedure which does not lead to sign reversals of the synaptic couplings during the learning mode.[27]

We would like to express our sincere gratitude to W. Kinzel for the inspiring climate without which this work would not have been conceived.

[1] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[2] W. A. Little and G. L. Shaw, Math. Biosci. **39**, 281 (1978).
[3] P. Peretto, Biol. Cybernet. **50**, 51 (1984).
[4] W. Kinzel, Z. Phys. B **60**, 205 (1985).
[5] L. Personnaz, I. Guyon, and G. Dreyfus, J. Phys. (Paris) Lett. **46**, L359 (1985).
[6] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).
[7] L. Personnaz, I. Guyon, G. Dreyfus, and G. Toulouse, to be published.
[8] A possible dependence on the synaptic connectivity J_{ij} itself is also implied.
[9] D. Hebb, *The Organization of Behaviour* (Wiley, New York, 1949).
[10] J. L. van Hemmen and R. Kühn, Phys. Rev. Lett. **57**, 913 (1986).
[11] H. Sompolinsky, to be published.
[12] J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, Nature **304**, 158 (1983).
[13] G. Parisi, J. Phys. A **19**, L617 (1986).
[14] J. P. Nadal, G. Toulouse, J. P. Changeux, and S. Dehaene, Europhys. Lett. **1**, 535 (1986).
[15] However, a first attempt to store correlated patterns with a local mechanism has been undertaken (see Ref. 4). The storing was achieved by simple replacement of all frustrated bonds by zero, which strongly limits the storing capacity, being only of order ln$N$.
[16] D. J. Amit, H. Gutfreund, and H. Sompolinsky, to be published.
[17] I. Kanter and H. Sompolinsky, Phys. Rev. A **35**, 380 (1987).
[18] Local stability means stability with respect to one spin flip.
[19] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).
[20] See also A. D. Bruce, A. Canning, B. Forrest, E. Gardner, and D. J. Wallace, University of Edinburgh Report No.

86/366, 1986 (to be published in Proceedings of the Conference on Neural Networks for Computing, Utah, 1986), a paper which we received in December, 1986, where the authors present a different type of Perceptron learning rule which leads to a symmetric synaptic matrix and does not use the threshold $T \sim 1$ which is important to our approach.

[21]For a Perceptron learning rule in a more general framework, see D. H. Ackley, G. E. Hinton, and T. J. Sejnowsky, Cognit. Sci. **9**, 147 (1985).

[22]By preliminary numerical simulations we found that the threshold condition $T=1$ leads to excellent retrieval properties of the system for patterns initially provided with noise even though the network is unsymmetric. With $T=0$ cycles appear. We note that the speed of convergence of the learning process also strongly depends on T.

[23]H. D. Block, Rev. Mod. Phys. **34**, 123 (1962).

[24]For random patterns the probability for a nontrivial solution to this system of equations becomes exponentially small for $N \rightarrow \infty$ as long as $p/N < 1$ (see a similar argument in Ref. 17).

[25]Details will be published elsewhere.

[26]B. Carnahan, H. A. Luther, and J. O. Wilkes, *Applied Numerical Methods* (Wiley, New York, 1969), p. 299.

[27]S. Diederich, M. Opper, R. D. Henkel, and W. Kinzel, in Proceedings of the Conference on Computer Simulation in Brain Science, Copenhagen, 20–22 August 1986, edited by E. Braun and A. Rhitle (Cambridge Univ. Press, to be published).