Vol. 26(1995) *ACTA PHYSICA POLONICA B* No 8

INTERACTION SPACE OF NEURAL NETWORKS WITH CLIPPED COUPLINGS

W. Tarkowski*

Physik–Department
der Technischen Universität München,
D–85747 Garching bei München, Germany

M. Lewenstein*

Joint Institute for Laboratory Astrophysics,
University of Colorado,
Box 440, Boulder, Colorado 80309, USA

(Received March 31, 1995)

Using replica approach we consider a neural network model with synaptic weights that are clipped from below and above by some numbers. Such a model is motivated by certain neurophysiological and technological arguments. For some values of coupling bounds, namely those that select binary synaptic efficacies approximately equal to ±1, a double-stage first-order phase transition is encountered. This transition occurs at zero temperature and leads to a change of the critical capacity value $\alpha_c(\kappa \rightarrow 0)$: $1.27 \rightarrow 0.83 \rightarrow 0$. An even more interesting situation is observed when synapses take only one value approximately (either -1 or $+1$). Here the storage ratio may grow from zero to some finite number if the couplings are allowed to fluctuate. We also study properties of the weights distribution and illustrate the contributions of Dirac's delta at both boundaries of the allowed range of the synaptic strengths.

PACS numbers: 87.10.+e, 05.50.+q, 64.60.Cn

One of the most important problems in the theory of neural networks concerns the storage capacity. For several years many attempts have been made in order to investigate properties of networks with synaptic couplings (weights, or strengths) which are described by a definite rule (such as, for instance, Hebb's rule, the pseudo-inverse rule, and many others — see Ref. [1]

---

* Permanent address: Centrum Fizyki Teoretycznej, Polska Akademia Nauk, 02-668 Warsaw, Poland.

for a detailed review). A new and powerful line of research in this field was introduced by Gardner [2], who proposed that one should look for the optimal coupling configuration in the space of all possible network weights.

Gardner's original idea has been widely extended since then (see, for example, Refs [3–11]). One of the main directions of subsequent investigations has been to impose some constraints on the synaptic couplings [3, 4, 12, 13]. In a recent series of papers [14–16] a network with the so–called sign–constrained synapses was considered. In such a system each coupling may take either positive or negative values (with an absolute value ranging from zero to infinity). For instance one can require that for a given i, $J_{ij} \leq 0$ for all j, *etc.* A network constructed in this way is of interest for biological, technological and cognitive reasons. One should, for example, evoke Dale's law [17], which, roughly speaking, says that in real neurophysiological systems there are actually two types of neurons: excitatory and inhibitory ones. Excitatory neurons connect to all others with positive synaptic strengths, so that for a fixed i, $J_{ji} \geq 0$ for all j. Conversely, inhibitory neurons connect only via negative couplings, *i.e.* for a fixed i, $J_{ji} \leq 0$ for all j. In general, the possibility of choosing of synaptic weights with arbitrary plus and minus signs may lead to unnecessary complexity of behaviour in the considered system (for further analysis of these problems — see Ref. [16]).

On the other hand, it is obvious that both in real situations (that concern neurophysiology) or in applications (which concern hardware, or information processing) couplings may attain only a limited range of values. For example, they have to be bounded from below and above. Such a case, which is the subject of the present paper, is interesting also from a purely physical point of view. In some conditions neural networks with synapses clipped from both below and above exhibit first-order phase transitions.

Let us consider a network with N binary neurons $\sigma_i = \pm 1$, where $i = 1, \ldots, N$. The dynamics of such a system in the absence of noise (*i.e.* at the temperature $T = 0$) takes the standard form of a simple, deterministic updating rule

$$\sigma_i(t + \Delta t) = \operatorname{sgn}\left[\sum_{j \neq i} J_{ij}\sigma_j(t)\right], \tag{1}$$

where $\{J_{ij}\}$ denotes the synaptic connections matrix, which in general does not need to be symmetric.

A class of $p = \alpha N$ (α designates here the storage ratio) specified states (patterns), ξ_j^μ, where the index $i = 1, \ldots, N$ enumerates the sites of the network and $\mu = 1, \ldots, p$ specifies the patterns, is said to be stored in the memory described by the dynamics (1), if the set $\left\{\xi_j^\mu\right\}$ defines its stationary states. The matrix $\{J_{ij}\}$ in such a case should satisfy for each μ and i the following inequalities

$$\xi_i^\mu \sum_{j\neq i} J_{ij}\xi_j^\mu > \kappa\sqrt{N}\,. \tag{2}$$

In addition, one usually assumes that the so-called spherical constraint

$$\sum_{j\neq i} J_{ij}^2 = N \tag{3}$$

must be fulfilled for every node i.

The stability parameter κ that enters expression (2) is a non-negative real number, which determines the size of domains of attraction for the stored configuration $\left\{\xi_j^\mu\right\}$. Of course, larger values of κ ensure stronger stabilization of the remembered data, but lead inevitably to a decrease in the storage capacity.

In order to put bounds on synaptic weight values, one requires that for each $i, j = 1, \ldots, N$ the following condition must be fulfilled

$$A \leq g_{ij}J_{ij} \leq B\,, \tag{4}$$

where A, B ($A \leq B$) are real numbers, and $g_{ij} = \pm 1$ have been introduced to control the sign of each coupling. In particular, when both A and B are positive (negative), g_{ij} is the sign (minus sign) of the corresponding coupling. To realize Dale's law one can use a separable form of the matrix g_{ij}, $g_{ij} = h_i g_j$ with $h_i, g_i = \pm 1$. We shall assume such a form in the following.

The probability that the solution of the inequalities (2) and (4) (with the constraint (3)) exists is, following Gardner, given by the fractional volume in the interaction space of the matrices $\{J_{ij}\}$

$$\begin{aligned} V \propto \prod_i \Bigg[\int \left(\prod_{j\neq i} \mathrm{d}J_{ij} \right) \delta\left(\sum_{j\neq i} J_{ij}^2 - N \right) \\ \times \prod_{j\neq i} \left[\Theta(g_{ij}J_{ij} - A)\,\Theta(B - g_{ij}J_{ij}) \right] \prod_\mu \Theta\left(\xi_i^\mu \sum_{j\neq i} \frac{J_{ij}}{\sqrt{N}} \xi_j^\mu - \kappa \right) \Bigg]\,, \end{aligned} \tag{5}$$

neglecting a constant normalization factor (see, for example, Ref. [2]) which plays no role in further considerations.

In this paper we restrict ourselves, for simplicity, to random, unbiased and uncorrelated patterns. Such data are characterized independently for each i and μ by the probability distribution

$$\Pr\left(\xi_j^\mu = \pm 1\right) = \tfrac{1}{2}\,. \tag{6}$$

Due to this fact, and thanks to the separability of the matrix g_{ij}, the theory is *locally gauge invariant* with respect to specific realizations of the quantities $\{g_{ij}\}$ (see Ref. [15]). After simple calculations one can show that the volume V is independent of the particular realizations of $\{g_{ij}\}$, and reads

$$V \propto \prod_i \left[\int_A^B \left(\prod_{j\neq i} \mathrm{d}J_{ij} \right) \delta\left(\sum_{j\neq i} J_{ij}^2 - N \right) \prod_\mu \Theta\left(\xi_i^\mu \sum_{j\neq i} \frac{J_{ij}}{\sqrt{N}} \xi_j^\mu - \kappa \right) \right]. \tag{7}$$

As usual in the theory of disordered systems, we are actually interested in calculating the mean value $\langle\!\langle \ln V \rangle\!\rangle$, averaged over all possible realizations of $\{\xi_i^\mu\}$. To evaluate the quantity $\langle\!\langle \ln V \rangle\!\rangle$, one makes use of the replica method, writing

$$\langle\!\langle \ln V \rangle\!\rangle = \lim_{n\to 0} \frac{\langle\!\langle V^n \rangle\!\rangle - 1}{n}. \tag{8}$$

We first take n to be natural, and perform an analytic continuation to $n = 0$. To calculate $\langle\!\langle V^n \rangle\!\rangle$, one introduces the order parameters

$$q^{\alpha\beta} = \frac{1}{N} \sum_{j\neq i} J_{ij}^\alpha J_{ij}^\beta \tag{9}$$

for $\alpha \neq \beta$, their conjugated counterparts $F^{\alpha\beta}$, and parameters E^α to assure the spherical normalization constraint (3). Note that $\alpha, \beta = 1, \ldots, n$ denote here the replica indices. The calculation of $\langle\!\langle V^n \rangle\!\rangle$ reduces, as usual, to an evaluation of the integral over the introduced order parameters. This integration can be performed using the saddle-point method [1, 2]. Assuming the replica-symmetric solution at the saddle point $q^{\alpha\beta} = q$, $F^{\alpha\beta} = F$, and $E^\alpha = E$ for each α and β, we obtain the expression for the expectation value of V^n in the form

$$\langle\!\langle V^n \rangle\!\rangle = C \int \mathrm{d}q\, \mathrm{d}F\, \mathrm{d}E\, e^{-Nn\mathcal{F}(q,F,E)}, \tag{10}$$

where C is a normalization constants, and the free-energy density $\mathcal{F}$ reads

$$\mathcal{F} = -\alpha G_1 - G_2 - \tfrac{1}{2} qF - \tfrac{1}{2} E. \tag{11}$$

The functions G_1, G_2 are given by

$$G_1 = \int \mathrm{D}t \ln H\left[\frac{\kappa + t\sqrt{q}}{\sqrt{1-q}} \right], \tag{12}$$

$$G_2 = \int \mathrm{D}t \ln\left\{ H\left[A\sqrt{E+F} + \frac{t\sqrt{F}}{\sqrt{E+F}} \right] - H\left[B\sqrt{E+F} + \frac{t\sqrt{F}}{\sqrt{E+F}} \right] \right\} + \tfrac{1}{2} \frac{F}{E+F} - \tfrac{1}{2} \ln(E+F), \tag{13}$$

with the Gaussian measure $\mathrm{D}t$ and the function $H(x)$

$$\mathrm{D}t = \frac{\mathrm{d}t}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} , \tag{14}$$

$$H(x) = \int_x^\infty \mathrm{D}t . \tag{15}$$

The critical capacity is obtained when the volume V shrinks to zero, which happens in the limit $q \to 1$. We anticipate that at criticality the order parameters behave as

$$F = \frac{\phi}{(1-q)^2} , \tag{16}$$

$$E = \frac{\psi}{(1-q)^2} , \tag{17}$$

$$E + F = \frac{\omega}{1-q} . \tag{18}$$

After some calculation one obtains the three saddle-point equations

$$\begin{aligned} 0 = \Bigg\{ \int \mathrm{D}t \Big[\Theta\Big(t + \omega \tfrac{A}{\sqrt{\phi}}\Big) + \Theta\Big(-t - \omega \tfrac{B}{\sqrt{\phi}}\Big) \Big] \Big[\Big(\tfrac{t}{\omega} + \tfrac{A}{\sqrt{\phi}} \Big) \Theta\Big(t + \omega \tfrac{A+B}{2\sqrt{\phi}} \Big) \\ + \Big(\tfrac{t}{\omega} + \tfrac{B}{\sqrt{\phi}} \Big) \Theta\Big(-t - \omega \tfrac{A+B}{2\sqrt{\phi}} \Big) \Big] t \Bigg\} - \tfrac{1}{\omega} + 1 , \end{aligned} \tag{19}$$

$$\begin{aligned} 0 = \Bigg\{ \int \mathrm{D}t \Big[\Theta\Big(t + \omega \tfrac{A}{\sqrt{\phi}}\Big) + \Theta\Big(-t - \omega \tfrac{B}{\sqrt{\phi}}\Big) \Big] \Big[\Big(A^2 - t^2 \tfrac{\phi}{\omega^2} \Big) \Theta\Big(t + \omega \tfrac{A+B}{2\sqrt{\phi}} \Big) \\ + \Big(B^2 - t^2 \tfrac{\phi}{\omega^2} \Big) \Theta\Big(-t - \omega \tfrac{A+B}{2\sqrt{\phi}} \Big) \Big] \Bigg\} + \tfrac{\phi}{\omega^2} - 1 , \end{aligned} \tag{20}$$

$$\phi = \alpha\, G(\kappa) , \tag{21}$$

where

$$G(\kappa) = \int \mathrm{D}t\, (t+\kappa)^2\, \Theta(t+\kappa) . \tag{22}$$

These equations may be solved numerically to obtain the critical capacity α_c. It is possible, however, to perform analytic calculations, and to show, for example, that if $A \to -\infty$ and $B = 0$ (or, alternatively, $A = 0$ and $B \to \infty$), $\alpha_c(\kappa \to 0) = 1$.

We stress that the above results have been obtained within the replica-symmetric theory. To be able to accept these results we must check the validity of the replica-symmetric Ansatz. To this end we examine the entropy function. Its value should tend to minus infinity at the saturation limit $q \to 1$. The most divergent term in the expression for the entropy function is proportional to $(1-q)^{-2}$. Depending on the sign of the factor which multiplies this term, the entropy may diverge either to plus or to minus infinity, if $q \to 1$. The regime of parameters in which replica-symmetric theory is valid can therefore be determined by imposing the requirement of the proper sign of the factor preceding the most divergent term of the entropy function.

After numerical calculations we have found that the replica-symmetric predictions are not valid if $-1 \le A \le 1$ and $-1 \le B \le 1$. The limiting case with $A = -1$, $B = 1$ is especially interesting. Here we have a network with binary couplings $J_{ij} = \pm 1$. All J_{ij} must then attain the absolute value 1, since otherwise the spherical constraint $\sum_{j \neq i} J_{ij}^2 = N$ cannot be fulfilled. The entropy function becomes strictly negative at the point $A = -1$, $B = 1$, which indicates that the replica-symmetric theory is no longer valid (note that for $A = -1$, $B = 1$ the couplings $\{J_{ij}\}$ are discrete and the entropy function may take only non-negative values). However, for $A = -1 - \varepsilon$, $B = 1 + \varepsilon$, with ε denoting a very small positive number, the replica-symmetric solution is correct and the entropy tends to minus infinity, if $q \to 1$. In this limit we obtain the critical capacity $\alpha_c(\kappa \to 0) = 1.27$, which is exactly equal to the number $4/\pi$ calculated by Gardner and Derrida [3] without replica-symmetry breaking for the case of binary weights $J_{ij} = \pm 1$. This result which obviously violates the bounds of informational capacity [1, 4], has been rejected as incorrect. We stress, however, that in the limit $A = -1 - \varepsilon$, $B = 1 + \varepsilon$, with $\varepsilon \to 0^+$, when couplings are very close to ± 1 (and only a very small fraction of them belongs to the set $[-1 + \varepsilon, 1 - \varepsilon]$), the solution $\alpha_c = 4/\pi$ is correct. As we see, even a very small relaxation of the requirement of binary coupling leads to enormous changes in capacity.

Note that for $A = -1$, $B = 1$, one should break the symmetry of replicas, and obtains then $\alpha_c \cong 0.83$ (see Ref. [4]). Furthermore, if $A = -1 + \varepsilon$, $B = 1 - \varepsilon$, the critical capacity α_c is zero (since the spherical constraint (3) can no longer be fulfilled). Therefore we have encountered here a double-stage (α_c: $1.27 \to 0.83 \to 0$) first-order phase transition, which occurs at zero temperature. This fact is of great importance: in all practical applications of networks with binary couplings (for example, in hardware, or information processing) it is impossible to assure that weights take *exactly* the values ± 1. Noise (of any origin) will inevitably cause the real coupling values to undergo some fluctuations around ± 1. In such a case the capacity of a system will be increased from 0.83 to 1.27. One should also notice that by taking $A = -1 - \varepsilon$, $B = 1 + \varepsilon$, we deal with a case that can be compared to the situation in which small thermal fluctuations of binary synaptic weights ± 1 occur. The limit $\varepsilon \to 0$ is then, in a sense, equivalent to the limit of vanishing fluctuations, or vanishing temperature

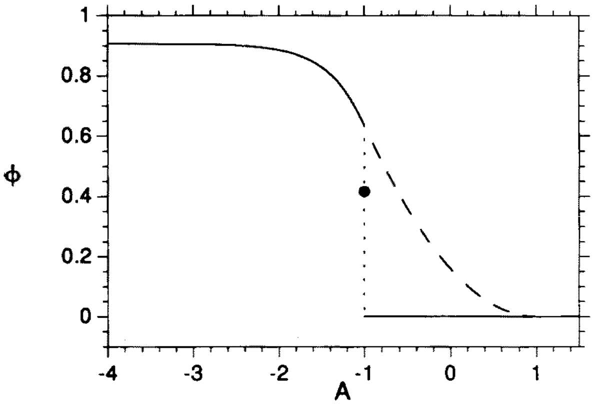

Fig. 1. The ϕ parameter as a function of variable A with $B = 1$. For $A = -1$ the double-stage first-order phase transition (α_c: $1.27 \rightarrow 0.83 \rightarrow 0$) is seen. Dashed line corresponds to the solution, for which replica-symmetric theory gives wrong results. Note that if $\kappa \rightarrow 0$, $\alpha_c = 2\phi$.

$T \rightarrow 0$. This situation is illustrated in Fig. 1.

An even more interesting case, however, is encountered for $A = -1 - \varepsilon$, $-1 < B < 1$ (or $B = 1 + \varepsilon$, and $-1 < A < 1$). Here, assuming that all g_{ij} have the same signs, we see that the coupling values are very close to either -1 or $+1$. Only a very small fraction of weights belongs now to the set $[-1 + \varepsilon, B - \varepsilon]$ (or $[A + \varepsilon, 1 - \varepsilon]$, respectively). It is worth stressing that in the limit $\varepsilon \rightarrow 0^+$ the storage capacity of a network remains finite and depends on the value of B (or A); on the other hand, strictly at the point $A = -1$, for $-1 < B < 1$ (or $B = 1$, for $-1 < A < 1$) the replica-symmetry must be broken, and $\alpha_c = 0$. Hence, again we conclude that even very small fluctuations of synaptic weights around -1 (or $+1$) can allow the system to store a number of patterns of the order of N. This phase transition is also interesting for technological applications. Note that learning requires here a choice of a few couplings only — most of them take the same value. In order to visualize these statements we have plotted (in Fig. 2) the dependence of the parameter ϕ (see Eq. (21)) on the variable A for several values of B. Dashed lines in both Figs 1 and 2 correspond to replica-symmetric predictions, which are not valid, if $-1 \leq A \leq 1$, $-1 \leq B \leq 1$.

The nature of replica-symmetry breaking in the above theory may be easily understood by simple physical (or rather geometrical) considerations (see Ref. [4]). For $A = -1 - \varepsilon$, $B = 1 + \varepsilon$ (and, similarly, if $A = -1 - \varepsilon$, $-1 < B < 1$, or $B = 1 + \varepsilon$, $-1 < A < 1$) the volume V (in situations not too far from saturation) forms a compact set **F** on the hypersphere $\sum_{j \neq i} J_{ij}^2 = N$, and the space of solutions of (2), (3), and (4) is connected; thus any

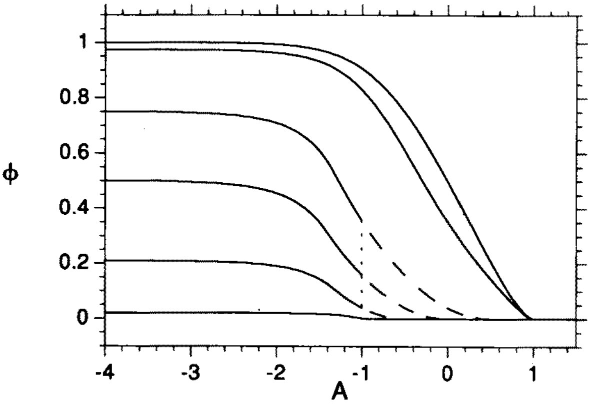

Fig. 2. The ϕ parameter as a function of the variable A for $B = 4$, 1.5, 0.5, 0, −0.5 and −0.9 (upper to lower). Dashed curves correspond to those values of A and B, for which replica-symmetric theory predictions are not valid. The first-order phase transition from a finite value of α_c to zero at $A = -1$ for $B = -0.9$, −0.5, 0 and 0.5 is marked by the dotted line. Note that for $\kappa \to 0$, $\alpha_c = 2\phi$.

solution may be continuously deformed into any other one. The critical capacity (for which only one connection matrix in the interaction space exists) is attained when $V \to 0$, and the typical overlap $q^{\alpha\beta}$ between distinct replicated configurations (which is equal to the cosine of the solid angle of the cone created by couplings that fulfil requirements (2), (3), and (4)) tends toward 1. For $\varepsilon = 0$, and hence if $J_{ij} = \pm 1$, all possible solutions (in the number of 2^N) are located on the hypercube, whose vertices belong to the hypersphere $\sum_{j \neq i} J_{ij}^2 = N$. In such a case the critical capacity behaviour (with the only one solution in the interaction space) may be thought of as equivalent to the situation in which the area of the surface of the set **F** on the hypersphere is around $1/2^N$. Such a surface contains typically only one point of the hypercube $J_{ij} = \pm 1$. From Gardner's original calculation, however, for $V \sim 1/2^N$ one gets $\alpha_c \approx 0.85$ and $q \sim 0.5$ (if $\kappa \to 0$). The value $q \sim 0.5$, which is obtained from these simple considerations, remains quite close to 0.56 given by the requirement that the zero-temperature entropy vanish. One step of replica-symmetry breaking gives the correct value of the critical capacity $\alpha_c \cong 0.83$ (see Ref. [4]).

Let us now examine the distribution of clipped synaptic efficacies [15], say the coupling $J_{i_0 j_0}$ for certain i_0 and j_0

$$
\begin{aligned}
\varrho(J) &= \langle\langle \delta(J - J_{i_0 j_0}) \rangle\rangle_{\{J_{ij}\}} \\
&\equiv \left\langle\!\!\left\langle \prod_i \left\{ \left[\int_{A \le g_{ij} J_{ij} \le B} \left(\prod_{j\neq i} \mathrm{d}J_{ij}\right) \delta(J - J_{i_0 j_0})\ \delta\Big(\sum_{j\neq i} J_{ij}^2 - N\Big) \right.\right.\right.\right. \\
&\quad \times \left. \prod_\mu \Theta\Big(\xi_i^\mu \sum_{j\neq i} \frac{J_{ij}}{\sqrt{N}} \xi_j^\mu - \kappa\Big) \right] \\
&\quad \times \left.\left.\left. \left[\int_{A \le g_{ij} J_{ij} \le B} \left(\prod_{j\neq i} \mathrm{d}J_{ij}\right) \delta\Big(\sum_{j\neq i} J_{ij}^2 - N\Big) \prod_\mu \Theta\Big(\xi_i^\mu \sum_{j\neq i} \frac{J_{ij}}{\sqrt{N}} \xi_j^\mu - \kappa\Big) \right]^{-1} \right\} \right\rangle\!\!\right\rangle_{\{\xi_i^\mu\}} \\
&= \left\langle\!\!\left\langle \lim_{n\to 0} \prod_i \left\{ \left[\int_{A \le g_{ij} J_{ij} \le B} \left(\prod_{j\neq i} \mathrm{d}J_{ij}\right) \delta(J - J_{i_0 j_0})\ \delta\Big(\sum_{j\neq i} J_{ij}^2 - N\Big) \right.\right.\right.\right. \\
&\quad \times \left. \prod_\mu \Theta\Big(\xi_i^\mu \sum_{j\neq i} \frac{J_{ij}}{\sqrt{N}} \xi_j^\mu - \kappa\Big) \right] \\
&\quad \times \left.\left.\left. \left[\prod_i \int_{A \le g_{ij} J_{ij} \le B} \left(\prod_{j\neq i} \mathrm{d}J_{ij}\right) \delta\Big(\sum_{j\neq i} J_{ij}^2 - N\Big) \prod_\mu \Theta\Big(\xi_i^\mu \sum_{j\neq i} \frac{J_{ij}}{\sqrt{N}} \xi_j^\mu - \kappa\Big) \right]^{n-1} \right\} \right\rangle\!\!\right\rangle_{\{\xi_i^\mu\}} .
\end{aligned}
\tag{23}
$$

Note that the resulting free energy density (and then the saddle-point value) has the same form as in Eq. (10).

After some calculations we obtain

$$
\begin{aligned}
\varrho(J) &= \Theta(gJ - A)\, \Theta(B - gJ)\, \frac{X}{\sqrt{2\pi}} \mathrm{e}^{-J^2 X^2/2} \\
&\quad + H(-gAX)\, \delta(gJ - A) + H(gBX)\, \delta(B - gJ)\,,
\end{aligned}
\tag{24}
$$

where $X = \omega/\sqrt{\phi}$. As we see the distribution $\varrho(J)$ contains two Dirac's delta contributions at both limits of the allowed range of synaptic values. If $A = -1$ and $-1 < B < 1$ (or $B = 1$, $-1 < A < 1$), however, $\varrho(J) = \delta(gJ - 1)$.

In Fig. 3 we have plotted the X parameter as a function of A for several values of B. Note that for $B > 1$ and $A \to 1$, $X \to \infty$.

Finally, we would like to add a comment about a proper learning algorithm, which can be easily constructed as follows. Let us start learning with all the couplings $\{J_{ij}\}$ taking values $(A + B)g_{ij}/2$. We then proceed exactly as is done in the perceptron algorithm (see Refs [2, 14, 18, 19]), with one exception: if the value of $g_{ij}J_{ij}$ after updating is lower than A or greater than B, then one puts the $g_{ij}J_{ij}$ equal to A or B, respectively. The convergence theorem for such an algorithm may be proved in a way very similar to the cases considered in Refs [2, 14, 18, 19].

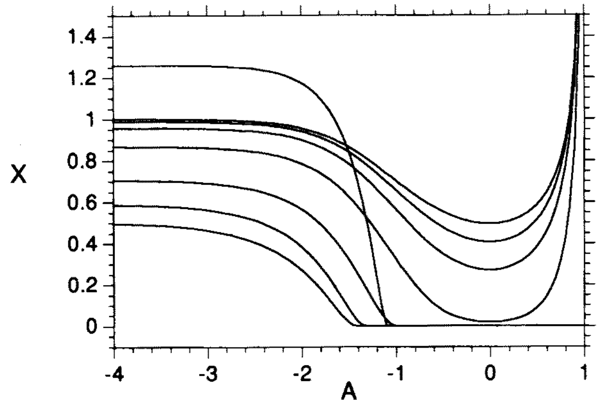

Fig. 3. The X parameter as a function of the variable A for $B = -0.9$, 4, 2.5, 2, 1.5, 1, -0.5 and 0 (upper to lower, looking at the right-hand side of the figure).

To summarize, we have investigated a network with clipped synaptic efficacies. The weights with large absolute values (for example, if $|J| > 2$) have a very small influence on the critical capacity. This is, of course, a simple consequence of the fact that only a very small fraction of couplings takes large values. Keeping the latter in mind, one may search for an efficient dilution algorithm and cut off of change (in a certain manner) some synapses that do not have a great influence on the storage capacity (*i.e.* those that take either the value zero or the large values, for instance).

In some special conditions, namely, if $A = -1$ and $B = 1$, we encounter a double-stage first-order phase transition (with "jumps" of the value of $\alpha_c(\kappa \to 0)$: $1.27 \to 0.83 \to 0$), which may play an important role in practical applications, enlarging the capacity of a network with binary couplings. An even more interesting situation occurs when synapses take only one value (either -1 or $+1$). Here the storage ratio may grow from zero to some finite number if one allows the couplings to fluctuate. We have also investigated the distribution of synaptic weights and found Dirac's delta contributions at both limits of the allowed range of efficacies values.

W.T. would like to thank *Alexander von Humboldt-Stiftung* for making possible his research stay in Germany. M.L. is grateful for the support and hospitality of JILA during his visiting fellowship. This work has been partially supported by the KBN Grants Nos. 2-2417-92-03 and 2-P302-045-06. W.T. acknowledges Dr Krzysztof Pachucki for his help in learning the *Mathematica* program.

Note

After completing this paper we learned about the recent article by R. W. Penney and D. Sherrington [20] on a similar subject. They considered a slightly less general case than ours, *i.e.* $|J| \leq B$.

REFERENCES

[1] D.J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks,* Cambridge University Press, Cambridge 1989.

[2] E. Gardner, *J. Phys. A* **21**, 257 (1988).

[3] E. Gardner, B. Derrida, *J. Phys. A* **21**, 271 (1988).

[4] W. Krauth, M. Mézard, *J. Phys. France* **50**, 3057 (1989).

[5] W. Tarkowski, M. Komarnicki, M. Lewenstein, *J. Phys. A* **24**, 4197 (1991).

[6] E. Barkai, D. Hansel, H. Sompolinsky, *Phys. Rev.* **A45**, 4146 (1992).

[7] H. S. Seung, H. Sompolinsky, N. Tishby, *Phys. Rev.* **A45**, 6056 (1992).

[8] A. Engel, H.M. Köhler, F. Tschepke, H. Vollmayr, A. Zippelius, *Phys. Rev.* **A45**, 7590 (1992).

[9] M. Griniasty, T. Grossman, *Phys. Rev.* **A45**, 8924 (1992).

[10] W. Tarkowski, M. Lewenstein, *J. Phys. A* **26**, 2453 (1993).

[11] W. Tarkowski, M. Lewenstein, *J. Phys. A* **26**, 3669 (1993).

[12] W. Krauth, M. Opper, *J. Phys. A* **22**, L519 (1989).

[13] H. Gutfreund, Y. Stein, *J. Phys. A* **23**, 2613 (1990).

[14] D.J. Amit, K.Y.M. Wong, C. Campbell, *J. Phys. A* **22**, 2039 (1989).

[15] D.J. Amit, C. Campbell, K.Y.M. Wong, *J. Phys. A* **22**, 4687 (1989).

[16] K.Y.M. Wong, C. Campbell, *J. Phys. A* **25**, 2227 (1992).

[17] J.C. Eccles, *The Physiology of Synapses*, Springer–Verlag, Berlin 1964.

[18] L.F. Abbott, T.B. Kepler, *J. Phys. A* **22**, L711 (1989).

[19] L.F. Abbott, *Network* **1**, 105 (1990).

[20] R.W. Penney, D. Sherrington, *J. Phys. A* **27**, 23 (1994).