The phase space of interactions in neural networks with definite symmetry

1989 J. Phys. A: Math. Gen. 22 1995

(http://iopscience.iop.org/0305-4470/22/12/005)

J. Phys. A: Math. Gen. **22** (1989) 1995-2008. Printed in the UK

The phase space of interactions in neural networks with definite symmetry

E Gardner, H Gutfreund and I Yekutieli

The Institute for Advanced Studies, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Received 26 January 1989

Abstract. We calculate the typical fraction of the phase space of interactions which solve the problem of storing a given set of p patterns represented as N-spin configurations, as a function of the storage ratio, $\alpha = p/N$, of the stability parameter, κ, and of the symmetry, η, of the interaction matrices. The calculation is performed for strongly diluted networks, where the connectivity of each spin, C, is of the order of $\ln N$. For each value of κ and η, there is a maximal value of α, above which the volume of solutions vanishes. For each value of κ and α, there is a typical value of η at which this volume is maximal. The analytical studies are supplemented by numerical simulations on fully connected and diluted networks, using specific learning algorithms.

Opening remark (by HG and IY)

The major part of this work was performed while the authors participated in the special programme, Spin Glasses and their Applications in Physics, Biology and Mathematics, at the Institute for Advanced Studies in Jerusalem. Elizabeth Gardner died a few months after she returned home from the Institute. We are grateful for having had the privilege to interact with her scientifically and socially. Elizabeth deserves credit for her fine contribution to this work, but the responsibility for any flaws in its present representation is entirely ours.

1. Introduction

The models of neural networks which have recently attracted so much attention are composed of N binary elements, represented by the spin variables $S_i = \pm 1$ and connected by the couplings J_{ij}. The dynamics of the network is, in the simplest case (zero noise and vanishing thresholds), given by

$$S_i(t+1) = \operatorname{sgn}\left(\sum_{j \neq i} J_{ij} S_j(t)\right). \tag{1}$$

The basic problem of these models is to find parameters J_{ij} for which a set of configurations $\{\xi_i^\mu\}$ $(i = 1, \ldots, N; \mu = 1, \ldots, p)$ are fixed points of the dynamics. There are two lines of approach to this problem. In the first approach, the J_{ij} are given by a specific storage prescription. Examples of this approach are what is called Hebb's rule, which is the basis of Hopfield's model (Hopfield 1982, Amit *et al* 1987) or the

0305-4470/89/121995+14$02.50

pseudo-inverse rule (Kohonen 1984), which has been applied to these models by Personnaz *et al* (1985) and studied analytically by Kanter and Sompolinsky (1987). In both cases the J_{ij} are symmetric, i.e. $J_{ij} = J_{ji}$. In the second approach, the parameters J_{ij} are found by a dynamical process of learning. An example of this approach is the perceptron algorithm (Minsky and Papert 1969), which was applied (Gardner 1988) to this class of models in order to find a set of J_{ij} which satisfy the inequalities

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

subject to the normalisation condition

$$\sum_{j \neq i} J_{ij}^2 = 1. \tag{3}$$

The patterns $\{\xi_i^\mu\}$ are fixed points of the dynamical equation (1) even if these inequalities are satisfied with $\kappa = 0$, but a finite κ is needed to ensure large basins of attraction. When $\kappa > 0$, it is necessary to specify the normalisation in order to avoid the freedom in the value of κ due to an overall scaling of the inequalities (2). A convergence theorem guarantees that a solution of these inequalities will be reached by the perceptron algorithm provided that a solution exists. The probability of the existence of such solutions can be deduced from calculating the fractional volume in the phase space of the interaction parameters J_{ij} in which (2) and (3) are satisfied. Such a calculation (Gardner 1988) gives a critical value $\alpha_c(\kappa)$ of $\alpha \equiv p/N$, such that this probability, in the limit $N \to \infty$, is equal to 1 for $\alpha \leqslant \alpha_c(\kappa)$ and equal to 0 otherwise. For random uncorrelated patterns $\alpha_c(0) = 2$.

In the present paper we investigate the symmetry properties of the matrices J_{ij} which satisfy (2) and (3). Let us define the symmetry parameter η as

$$\eta = \frac{\sum_{i \neq j} J_{ij} J_{ji}}{\sum_{i \neq j} J_{ij}^2}. \tag{4}$$

An equivalent definition of η is

$$\eta = \frac{\sum_{i \neq j} [(J_{ij}^s)^2 - (J_{ij}^a)^2]}{\sum_{i \neq j} J_{ij}^2} \tag{5}$$

where $J_{ij}^{a,s} = \frac{1}{2}(J_{ij} \pm J_{ji})$ are the symmetric and antisymmetric components of J_{ij}. When $\eta = 1$ the matrix is symmetric and when $\eta = -1$ it is fully antisymmetric. When $\eta = 0$, there is no correlation between J_{ij} and J_{ji} on the average, which also means that the symmetric and antisymmetric components have equal weights. The symmetry of the matrix J_{ij} can also be characterised by the ratio, k, of the antisymmetric and symmetric components. This parameter is related to η by

$$\eta = \frac{1 - k^2}{1 + k^2}. \tag{6}$$

The relevance of symmetry in neural network models has been pointed out by several authors. Krauth *et al* (1988) have shown recently that the basins of attraction depend not only on κ, but also on η. For a given κ and $\alpha < \alpha_c(\kappa)$, solutions with smaller values of η (less symmetric) have larger basins of attraction. The role of asymmetry was also emphasised, in a different context, by Parisi (1986). In a symmetrically coupled network there are many (exponentially many) fixed points outside of the basins of attraction of the stored patterns. Parisi suggested that a distinction be

made between a 'meaningful' operation of the network, when the system converges to one of the learned patterns in response to a 'known' external stimulus (initial configuration), and a 'meaningless' behaviour, when the system wanders around in the configuration space until a 'known' stimulus appears. To make this behaviour plausible, one has to eliminate the spurious fixed points, and this can be done by increasing the asymmetry. Such a relation between the asymmetry and the 'landscape' was studied in the case of spin glasses (with deterministic dynamics) by Gutfreund *et al* (1988). The effect of the symmetry of the J_{ij} on the dynamics in spin-glass systems was investigated by Hertz *et al* (1987) and by Crisanti and Sompolinsky (1988).

The preceding paragraph motivates the study of the symmetry properties of the solutions of (2) and (3). In particular, one can ask the following questions. What are the typical values of η and what is their distribution in the subspace of matrices J_{ij} which solve (2) and (3) for given α and κ? What is the storage capacity for a fixed symmetry, i.e. what is $\alpha_c(\kappa, \eta)$? What values of η are reached by specific learning algorithms, and what is their dependence on the initial J_{ij}? These and related questions are the subject of the present paper.

In § 2 we extend the calculation of Gardner (1988) to derive the fractional volume of solutions (2) and (3) with a specific symmetry η. It turns out that this can be done analytically only in a diluted network in which each neuron is connected, on the average, to $C \simeq \ln N$ other neurons. As $N \to \infty$, the volume is sharply peaked around a typical $\eta(\alpha, \kappa)$. In § 3 we investigate the storage capacity as a function of η. In § 4 we present results of numerical algorithms for finding solutions in diluted and fully connected networks. The main conclusion is that typical solutions in fully connected networks are significantly more symmetric than in diluted networks. We also investigate the accessibility, via the learning algorithms, of solutions with different values of η.

2. The fractional volume of solutions with a given symmetry

The fractional volume, V, in the space of the J_{ij}, of solutions of the inequalities (2) with a definite value of the symmetry parameter η is given by

$$\prod_i \left[\int \prod_{j \neq i} \mathrm{d}J_{ij} \prod_\mu \theta\left(\xi_i^\mu \sum_{j \neq i} \frac{J_{ij}}{\sqrt{N}} \xi_j^\mu - \kappa \right) \delta\left(\sum_{j \neq i} J_{ij}^2 - N \right) \delta\left(\sum_{j \neq i} J_{ij} J_{ji} - \eta N \right) \right] \times \left\{ \prod_i \left[\int \prod_{j \neq i} \mathrm{d}J_{ij} \delta\left(\sum_{j \neq i} J_{ij}^2 - N \right) \right] \right\}^{-1} \tag{7}$$

where $\theta(x)$ is the step function: $\theta(x) = 1$ when $x > 0$, and $\theta(x) = 0$ otherwise. In the absence of restrictions on the correlations between J_{ij} and J_{ji}, i.e. without the second δ function in the first term in large square brackets, the sites i are decoupled and the fractional volume can be calculated for each of them separately. The typical fractional volume for all possible representations of the $\{\xi_i^\mu\}$ requires the averaging of $\ln V$ over the probability distribution of the ξ_i^μ and can be done using the replica method (Gardner 1988). This calculation cannot be simply extended to the present case, because now the different rows of the matrix J_{ij} are coupled. The difficulty arises in the averaging over the ξ_i^μ. In the replica method this averaging is performed over V^n, which is obtained from (7) by adding a replica index, α, to each of the variables J_{ij} and taking a product over α. Using the integral representation of the θ functions one can average

over the probability distribution of the ξ_i^μ:

$$\left\langle \prod_{\alpha,\mu,i} \theta\left(\xi_i^\mu \sum_{j\neq i} \frac{J_{ij}^\alpha}{\sqrt{N}} \xi_j^\mu - \kappa\right)\right\rangle_\xi$$
$$= \int_\kappa^\infty \prod_{\alpha,\mu,i} d\lambda_{\mu i}^\alpha \int_{-\infty}^{\infty} \prod_{\alpha,\mu,i} \frac{dx_{\mu i}^\alpha}{2\pi} \exp\left(\mathrm{i} \sum_{\alpha,\mu,i} x_{\mu i}^\alpha \lambda_{\mu i}^\alpha\right)$$
$$\times \left\langle \exp\left(-\mathrm{i} \sum_{\alpha,\mu,i} x_{\mu i}^\alpha \xi_i^\mu \sum_{j\neq i} \frac{J_{ij}^\alpha}{\sqrt{N}} \xi_j^\mu\right)\right\rangle_\xi . \tag{8}$$

Performing the average over the ξ_i^μ, we get a cumulant expansion which does not converge, because of the summation over i (see appendix 1).

However, this calculation can be performed in a symmetrically diluted network in which each site is connected, on the average, to C other sites, so that the existence of a bond from i to j implies also the existence of a bond from j to i. Note that this is different from the diluted model of Gardner *et al* (1987) in which the existence of a bond (i,j) does not depend on the existence of a bond (j,i). Continuing with our diluted model, the normalisation condition is modified to

$$\sum_{j\neq i} J_{ij}^2 = C. \tag{9}$$

N is replaced by C in (7) and (8), and the summations over j in these equations are on the C sites connected to site i. When $C \simeq \ln N$, the higher terms in the cumulant expansion are of lower order of magnitude than the leading quadratic term (appendix 1). The result of the averaging over the ξ_i^μ is

$$\left\langle \exp\left(-\mathrm{i} \sum_{\alpha,\mu,i} x_{\mu i}^\alpha \xi_i^\mu \sum_{j\neq i} \frac{J_{ij}^\alpha}{\sqrt{C}} \xi_j^\mu\right)\right\rangle_\xi$$
$$= \prod_\mu \exp\left(-\frac{1}{2} \sum_{\alpha,\beta,i} x_{\mu i}^\alpha x_{\mu i}^\beta \sum_{j\neq i} \frac{J_{ij}^\alpha J_{ij}^\beta}{C} - \frac{1}{2} \sum_{\alpha,i} x_{\mu i}^\alpha \sum_{\beta,j\neq i} x_{\mu j}^\beta \frac{J_{ij}^\alpha J_{ji}^\beta}{C}\right). \tag{10}$$

The first term in the exponential suggests that we define the local order parameter

$$q_{\alpha b}^i = \frac{1}{C} \sum_{j\neq i} J_{ij}^\alpha J_{ij}^\beta . \tag{11}$$

To treat the second term, we define a parameter $g_{\mu i}^{\alpha\beta}$

$$g_{\alpha\beta}^{\mu i} = \sum_{j\neq i} x_{\mu j}^\beta \frac{J_{ij}^\alpha J_{ji}^\beta}{C}.$$

Assuming that g is independent of the site i, summing over i and dividing by N, we have

$$g_{\alpha\beta}^\mu = \frac{1}{N} \sum_j x_{\mu j}^\beta h_{\beta\alpha}^j$$

where we have defined a new local order parameter

$$h_{\alpha\beta}^i = \frac{1}{C} \sum_{j\neq i} J_{ij}^\alpha J_{ji}^\beta . \tag{12}$$

In appendix 2 we calculate $\langle V^n \rangle$ for replica and site-symmetric solutions. We did not check the stability of the solutions to replica symmetry breaking. When there are

no restrictions on the correlations between J_{ij} and J_{ji}, the replica-symmetric solutions are stable for all α and κ (Gardner 1988). We believe that in the present case they are stable at least for a range of values of η within which the volume of solutions attains its maximum. However, as will be discussed in § 3, there are signs of replica symmetry breaking for values of η in an interval $-1 < \eta < \eta_m(\kappa)$, which increases with κ. The assumption of site symmetry can be viewed as a mean-field ansatz. We have checked it by simulations, calculating the parameters $q^i_{\alpha\beta}$ and $h^i_{\alpha\beta}$ for different solutions obtained by the perceptron algorithm. The width of the distributions of these parameters over the sites i decreases sharply as N increases, and it seems that in the $N \to \infty$ limit, their distributions will become δ functions.

The result for the typical volume is

$$V_t(\alpha, \kappa, \eta) = \exp\left\{ CN\left(\tfrac{1}{2}\ln(1-q) + \tfrac{1}{4}\ln(1-x^2) + \frac{1}{2}\frac{q - hx}{(1-q)(1-x^2)}\right)\right\}$$
$$\times \exp\left\{ CN\left[-\frac{1}{2}\frac{\alpha r^2}{(1-q)x} + \alpha \int \mathrm{D}t \ln H\left(\frac{\sqrt{q}\, t + \kappa - r}{\sqrt{1-q}}\right)\right]\right\} \tag{13}$$

where

$$x = (\eta - h)/(1-q) \tag{14}$$

$$H(z) = \int_z^\infty \mathrm{D}\lambda \tag{15}$$

$$\mathrm{D}\lambda = \frac{\mathrm{d}\lambda}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}\lambda^2). \tag{16}$$

The variable r is an additional parameter which appears in the calculation (see appendix 2). The saddle-point equations for q, h and r are

$$\frac{q - 2hx + qx^2}{(1-x^2)^2} = \alpha F_2(\kappa - r, q) \tag{17}$$

$$\frac{h - 2qx + hx^2}{(1-x^2)^2} = \alpha (F_1(\kappa - r, q))^2 \tag{18}$$

$$r = xF_1(\kappa - r, q) \tag{19}$$

where

$$F_k(\kappa - r, q) = \int_{-\infty}^{\infty} \mathrm{D}t \left(\frac{\sqrt{1-q}\exp[-(\sqrt{q}\, t + \kappa - r)^2/2(1-q)]}{\sqrt{2\pi}\, H((\sqrt{q}\, t + \kappa - r)/\sqrt{1-q})}\right)^k. \tag{20}$$

One can solve equations (17)-(19) for given values of α, κ and η, and substitute the saddle-point values of q, h and r into (13) to obtain the typical volume. In figures 1(a) and 1(b), we plot the typical volume per bond, normalised to unity at $\alpha = 0$, $\eta = 0$, as a function of η for several values of α, and for $\kappa = 0$ and $\kappa = 0.5$, respectively. It should be emphasised that these curves represent the volume per bond. To obtain the relative volume in the space of matrices J_{ij}, one has to raise these curves to the (CN)th power. Thus, when $N \to \infty$ the relative volume becomes sharply peaked around a typical $\eta(\alpha, \kappa)$.

We observe that the volume vanishes when $\eta \to \pm 1$ (if it has not vanished for lower $|\eta|$). In the case $\eta = -1$, corresponding to an antisymmetric J_{ij}, there are no solutions

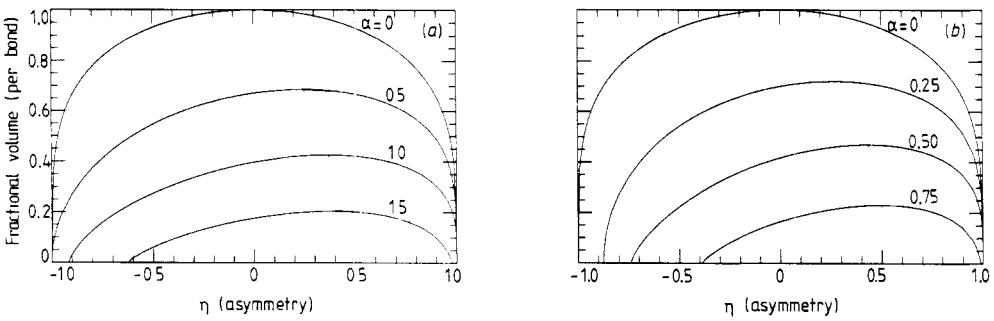

Figure 1. The typical fractional volume (per bond) plotted against η for (*a*) $\kappa = 0$ and (*b*) $\kappa = 0.5$ and several values of α. The dotted curve connects points of maximal volume.

to (2). If one assumes that such a solution exists, one immediately arrives at a contradiction, because the sum of the left-hand side of (2) over i is identically zero when $J_{ij} = -J_{ji}$. On the other hand, when $\eta = 1$, corresponding to a symmetric J_{ij}, such solutions exist, but they are confined to a region in a subspace of lower dimension than that of the space of all the J_{ij} ($\frac{1}{2}CN$ as against CN) so that their fractional volume vanishes. These limits are discussed in the next section.

To obtain the values of η at which the volume reaches its maximum, note that V_t depends on η only through x. Thus, the maximum of V_t with respect to η is its maximum with respect to x. The saddle-point equation for h at the maximum reduces to

$$-\frac{1}{2}\frac{x}{(1-q)(1-x^2)} = 0. \tag{21}$$

Hence the maximum of V_t is obtained at $x = 0$, which implies $h = \eta$ and $r = 0$. Solving (17) and (18) with these values, one finds the asymmetry parameter η at the maximum, which turns out to be positive for all κ. The maxima for different values of α are connected by the dotted curves in figures 1(*a*) and 1(*b*). It is interesting to point out that at the maximum the expression in (13) reduces to $(V_t^0)^{CN}$, where V_t^0 is the typical volume per bond when no correlations between J_{ij} and J_{ji} are imposed (Gardner 1988).

3. The critical storage capacity

The critical storage capacity is obtained when the fractional volume shrinks to zero. This happens when there is only one solution, i.e. when $q \to 1$. The second logarithmic term of (13) indicates that $|x| \leqslant 1$, so that by the definition of x (equation (14)), $q \to 1$ implies that the limit $h \to \eta$ has to be taken as well. In the limit $q \to 1$, the functions F_1, F_2, defined in (20), become

$$F_k(\kappa - r, q \to 1) = \int_{-\kappa + r}^{\infty} \mathrm{D}t(t + \kappa - r)^k. \tag{22}$$

Thus, to obtain the critical storage capacity, $\alpha_c(\kappa, \eta)$, at a given η, we have to solve (17)-(19), where on the left-hand side we put $q = 1$, $h = \eta$, and on the right-hand side we use (22). In figure 2, we plot $\alpha_c(\kappa, \eta)$ for several values of κ. Two points deserve special attention. First, the $\kappa = 0$ curve indicates the existence of solutions with $\eta = -1$,

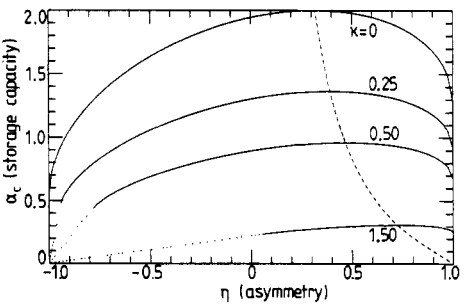

Figure 2. The critical storage capacity plotted against η, for $\kappa = 0$, 0.25, 0.5 and 1.5. The solid curve represents critical solutions with $q = 1$, and dotted curves represent solutions with $0 < q < 1$. The broken curve connects points of maximal volume.

although we showed that such solutions do not exist. This apparent contradiction will be discussed below. Second, for every $\kappa > 0$ there exists a minimal value of η below which there are no solutions with $q = 1$. These are the values of η at which the solid curves turn into dotted curves. The meaning of the dotted curves will also be discussed below.

Just like the maximum of the volume, the maximum of $\alpha_c(\kappa, \eta)$ with respect to η is also obtained at $x = 0$ and $r = 0$. This immediately gives the familiar result (Gardner 1988)

$$\alpha_c(\kappa) = \left(\int_{-\kappa}^{\infty} Dt(t+\kappa)^2 \right)^{-1}. \tag{23}$$

The value of η at the maximum is found from (17) (with $x = 0$, $\alpha = \alpha_c$) to be

$$\eta = \left(\int_{-\kappa}^{\infty} Dt(t+\kappa) \right)^2 \left(\int_{-\kappa}^{\infty} Dt(t+\kappa)^2 \right)^{-1}. \tag{24}$$

For example, at $\kappa = 0$ when $\alpha_c = 2$, the symmetry parameter is $\eta = 1/\pi$. Thus, the best storage capacity is obtained when J_{ij} and J_{ji} are correlated. If we insist on solutions with no such correlations ($\eta = 0$), we find that $\alpha_c(0, 0) = 1.94$. The maxima of $\alpha_c(\kappa, \eta)$ are indicated by the broken curve in figure 2.

In order to find the critical storage capacity for symmetric matrices, note that in the limit $\eta = 1$ by definition $h = q$ implying that $x = 1$. From (17)-(19), one finds in this limit

$$\alpha_c(\kappa, 1) = \frac{1}{2} \left(\int_{-\kappa+r}^{\infty} Dt(t+\kappa-r)^2 + r^2 \right)^{-1} \tag{25}$$

where r is found by solving

$$r = \int_{-\kappa+r}^{\infty} Dt(t+\kappa-r). \tag{26}$$

For $\kappa = 0$, $\alpha_c(0, 1) = 1.278$.

Let us now discuss the other end of values of η. We have seen that $x = 1$ when $\eta = 1$. One can deduce, in the same manner, that $x = -1$ when $\eta = -1$. Between these two limits x increases monotonically with η. It can be seen from (19), in the limit $q \to 1$, that r increases monotonically with x, from $r = -\infty$ up to a positive value determined by (19) with $x = 1$. Therefore, η is a monotonically increasing function of

r, and it attains its lowest value in the limit $r \to -\infty$. Taking this limit one finds the 'most antisymmetric' solutions

$$\eta(r \to -\infty) = \frac{\kappa^2 - 2}{\kappa^2 + 2} \equiv \eta_m(\kappa) \tag{27}$$

$$\alpha_c(r \to -\infty) = \frac{1}{\kappa^2 + 2}. \tag{28}$$

For each value of κ there exists a minimal value of η for which a critical solution (with $q = 1$) exists. These are the left end-points of the solid curves in figure 2. Note that when $\kappa > \sqrt{2}$, there are no such solutions with negative η. For $\kappa = 0$, the minimal value of η is -1, and for $\kappa > 0$, it is always greater than -1. The existence of solutions at $\kappa = 0$ and $\eta = -1$ can be understood when one considers the distribution of the local 'fields' $h_i^\mu = \xi_i^\mu \Sigma_{j \neq i} J_{ij} \xi_j^\mu$. This distribution has a δ function component at κ and a Gaussian tail above it. When $\kappa \to 0$ and $\eta \to -1$, this distribution reduces to a δ function at $\kappa = 0$. Thus, in this limit, the inequalities (2) become strict equalities.

Note that the conclusion that solutions with the lowest value of η correspond to $r \to -\infty$ is not restricted to $q = 1$. For any given value of q there exists a 'most antisymmetric' solution, given by $r \to -\infty$. In this limit

$$F_1(\kappa - r, q) \sim \kappa - r \tag{29}$$

$$F_2(\kappa - r, q) \sim (\kappa - r)^2 + q \tag{30}$$

leading to

$$\eta(r \to -\infty) = \frac{(2q - 1)\kappa^2 - 2q}{\kappa^2 + 2q} \tag{31}$$

$$\alpha_c(r \to -\infty) = \frac{q}{\kappa^2 + 2q} \tag{32}$$

which are generalisations of (27) and (28) for $q < 1$. For $\kappa > 0$, we can eliminate q from (31) and (32) to get the maximal storage capacity for highly antisymmetric solutions (with $q < 1$)

$$\alpha_c(\kappa, \eta) = (1 + \eta)/2\kappa^2. \tag{33}$$

This is the critical storage capacity for $\eta < \eta_m(\kappa)$, where η_m is defined in (27). It is represented by the dotted curves in figure 2.

The maximal α is always obtained when the volume of solutions shrinks to zero. If this volume is convex, then there remains in this limit a single solution, therefore $q = 1$. In the range $-1 < \eta < \eta_m(\kappa)$, the volume of solutions with a given η shrinks to zero, but $q < 1$. This means that there are two, or more, isolated solutions such that any interpolation between them gives solutions with higher η. Thus, in this range the volume of solutions breaks up into separate regions in the space of the J_{ij}. This is a sign of replica symmetry breaking.

4. Numerical simulations

In the present section we study the symmetry of solutions obtained by a specific learning algorithm. The algorithm proceeds as follows. At each site one defines (consecutively

for every pattern) the parameters

$$\varepsilon_i^\mu = \theta(\|J\|_i \kappa - h_i^\mu) \tag{34}$$

where

$$h_i^\mu = \xi_i^\mu \sum_{j \neq i} J_{ij} \xi_j^\mu \tag{35}$$

$$\|J\|_i = \left(\sum_{j \neq i} J_{ij}^2 \right)^{1/2}. \tag{36}$$

Starting from some initial matrix J_{ij}^0, the current values of J_{ij} are modified by

$$\Delta J_{ij} = \varepsilon_i^\mu \frac{1}{N} f(h_i^\mu) \xi_i^\mu \xi_j^\mu \tag{37}$$

until all the ε_i^μ are zero. We shall present results for two choices of the function $f(h_i^\mu)$.

(i) The perceptron algorithm

$$f(h_i^\mu) = \lambda \tag{38}$$

is guaranteed to converge to a solution, if one exists, in a finite number of steps.

(ii) The relaxation algorithm

$$f(h_i^\mu) = \lambda(\kappa - h_i^\mu)\|J\|_i \tag{39}$$

is a modification of an algorithm suggested recently by Abbott and Kepler (1988). This algorithm converges (if a solution exists) for $0 < \lambda \leqslant 2$, and is most efficient for $\lambda = 2$.

We have seen in § 3 that the fractional volume of solutions is sharply peaked (as $N \to \infty$) around a typical symmetry. The question is whether the broad range of values of η of existing solutions, around this typical value, but having a vanishingly small relative volume, can ever be reached by the learning algorithms. This can be done to some extent, when the search in the space of solutions is started with an initial matrix with a specific symmetry. To demonstrate this, we first present results of simulations for a fully connected network using the perceptron algorithm. We are particularly interested in the effect of the starting matrix and of the 'learning' step λ on the final symmetry.

In figure 3(a) we show results for the symmetry of the solutions when the initial $\{J_{ij}^0\}$ is a random Gaussian matrix with specified values of $\eta_0 = -1, 0, 1$. The size of the network is $N = 50$, the learning step is $\lambda = 1$, and the stability parameter is $\kappa = 0$. We wish to point out at the outset that going to larger networks does not change the results described below. One observes that the 'memory' of the initial symmetry decreases when α increases and is lost completely for $\alpha > 1$. Figure 3(b) shows that this 'memory' is preserved to higher values of α when $\lambda < 1$, while the dependence on the initial conditions is lost faster when $\lambda > 1$ (figure 3(c)). This dependence on λ can be easily understood in the case of $\kappa = 0$. In this case the normalisation of the J_{ij} plays no role and can be omitted. The final J_{ij} has the form

$$J_{ij} = J_{ij}^0 + \lambda \sum_\mu n_i^\mu \xi_i^\mu \xi_j^\mu \tag{40}$$

where n_i^μ is the number of times that the inequality (2) was not satisfied by pattern μ at site i. For large λ, the second term is dominant and the symmetry of J_{ij} is determined mainly by the distribution of the parameters n_i^μ. For small λ the algorithm performs a fine search for solutions, as close as possible to the initial matrix J_{ij}^0. This explains the considerably stronger dependence of η on its initial value in figure 3(a).

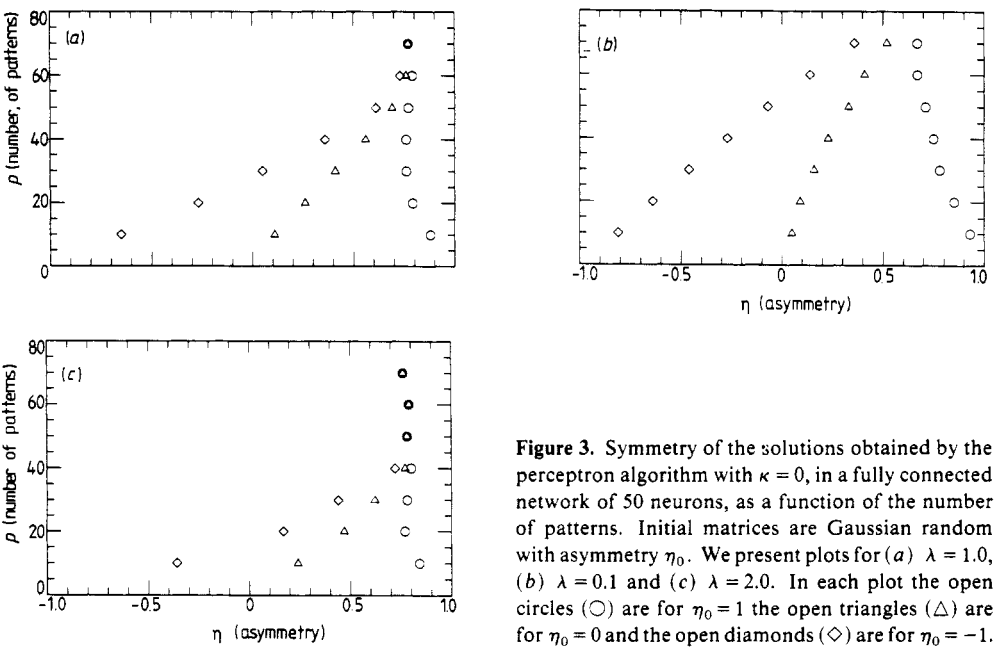

Figure 3. Symmetry of the solutions obtained by the perceptron algorithm with $\kappa = 0$, in a fully connected network of 50 neurons, as a function of the number of patterns. Initial matrices are Gaussian random with asymmetry η_0. We present plots for (a) $\lambda = 1.0$, (b) $\lambda = 0.1$ and (c) $\lambda = 2.0$. In each plot the open circles (○) are for $\eta_0 = 1$ the open triangles (△) are for $\eta_0 = 0$ and the open diamonds (◇) are for $\eta_0 = -1$.

It is very hard to continue these graphs to higher α, for two reasons.

(i) The probability that a solution exists for a given set of $\{\xi_i^\mu\}$ and $\alpha < 2$, approaches 1 when $N \to \infty$ (Venkatesh 1986), but for finite N the corrections may be non-negligible. For example, for $N = 50$ and $\alpha = 1.8$, this probability is about 0.001.

(ii) The time required to find a solution, if it exists, shows a critical slowing down (Opper 1988).

It is reasonable to assume that the value of η obtained in figures 3(a) and 3(c) for the highest α, when all memory of the initial conditions has been lost, is close to the most probable value of η. Judging from the weak dependence of η at the maximum fractional volume for a dilute network, when $\alpha > 1$ (see figure 1(a)), we can also assume that this value of η will not change much up to α_c, where it is determined uniquely. Thus, we conclude that $\eta_c(0) \simeq 0.7$–0.8, which shows that the symmetry at α_c in a fully-connected network is significantly higher than in a dilute network, where we found analytically $\eta_c(0) = 1/\pi$.

In figure 4(a) we use the relaxation algorithm with $\lambda = 2$ to obtain the symmetry of the final solution in a dilute network. It is hard to satisfy the conditions for strong dilution ($C \simeq \ln N$) in simulations, because C cannot be very small if one wants to find solutions for $\alpha = p/C > 1$. Nevertheless, even for $N = 500$ and $C = 50$ one observes a significant reduction of the symmetry of solutions at high α, compared to the fully connected network, to values close to what is expected at strong dilution. The continuous curve represents the symmetry of the typical solutions in a strongly diluted network. For comparison, we present in figure 4(b) results for a fully connected network of $N = 50$, using the same algorithm. The symmetry at the highest accessible α is roughly the same as in figure 3(a), although for lower α this algorithm shows a weaker dependence on the initial conditions than the perceptron algorithm with $\lambda = 1$.

In this section we have presented results for $\kappa = 0$. One finds similar behaviour at finite κ, except that in this case the typical solutions are more symmetric, as expected.

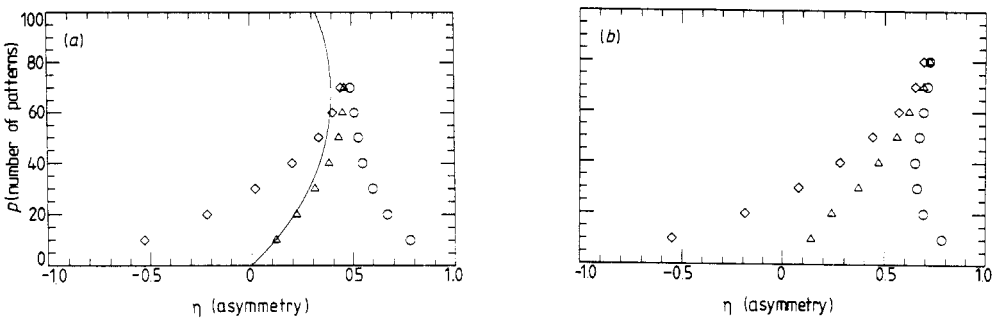

Figure 4. Solutions with the relaxation algorithm (with $\lambda = 2$) for (a) diluted networks with $N = 500$ and $C = 50$ and (b) fully connected networks with $N = 50$. Initial matrices are Gaussian random with asymmetry $\eta_0 = 1$ (○), $\eta_0 = 0$ (△) or $\eta_0 = -1$ (◇).

5. Concluding remarks

We have calculated the fractional volume of matrices J_{ij}, of a given symmetry, for which $p = \alpha N$ random patterns are stable fixed points of the network dynamics. This could be done analytically only in strongly diluted networks. One novel feature which emerges in the calculation is that the matrices J_{ij} at maximal storage in the range $-1 < \eta < \eta_m(\kappa)$ are not characterised by $q = 1$. This indicates that when critical storage is approached the space of solutions of (2) and (3) in this range is not simply connected and that, probably, replica symmetry is broken. This observation calls for a more thorough study of the space of solutions, in particular the stability of the replica symmetric solution and how its breaking appears as a function of α, κ and η.

In the introduction we have mentioned the study of the symmetry of the J_{ij} by its suggested effect on the performance of such networks as models for associative memory. In the present paper we have only studied the existence of solutions of definite symmetry and how they can be reached by specific learning algorithms. We have not investigated the effect of the symmetry on the stability of the learned patterns to noise, their basins of attraction, and related questions. This is a subject of a separate study.

Acknowledgment

We are grateful to the staff of the Institute for Advanced Studies for their hospitality. We benefited from numerous discussions on questions addressed in this work with our colleagues M Mézard, P Peretto, H Sompolinsky and G Toulouse.

Appendix 1

In this appendix we show that the average of

$$\exp\left(-\mathrm{i}\sum_{\alpha,\mu,i} x^{\alpha}_{\mu i}\xi^{\mu}_{i}\sum_{j\neq i}\frac{J^{\alpha}_{ij}}{\sqrt{C}}\xi^{\mu}_{j}\right) \tag{A1.1}$$

over the probability distribution

$$P(\xi^{\mu}_{i}) = \tfrac{1}{2}\delta(\xi^{\mu}_{i}-1)+\tfrac{1}{2}\delta(\xi^{\mu}_{i}+1) \tag{A1.2}$$

gives (10). We start with the cumulant expansion

$$\langle e^z \rangle = \exp\left(\sum_{k=1}^{\infty} C_k \right) \tag{A1.3}$$

where z is the argument in (A1.1). The first four cumulants in this expansion are

$$\begin{aligned} C_1 &= \langle z \rangle \\ C_2 &= \tfrac{1}{2}(\langle z^2 \rangle - \langle z \rangle^2) \\ C_3 &= \tfrac{1}{6}(\langle z^3 \rangle - 3\langle z^2 \rangle \langle z \rangle + 2\langle z \rangle^3) \\ C_4 &= \tfrac{1}{24}(\langle z^4 \rangle - 4\langle z^3 \rangle \langle z \rangle - 3\langle z^2 \rangle^2 + 12\langle z^2 \rangle \langle z \rangle^2 - 6\langle z \rangle^4). \end{aligned} \tag{A1.4}$$

The first cumulant clearly vanishes, which greatly simplifies the subsequent terms. The evaluation of the kth moment, $\langle z^k \rangle$, involves the average

$$\langle \xi_{i_1}^{\mu_1} \xi_{j_1}^{\mu_1} \xi_{i_2}^{\mu_2} \xi_{j_2}^{\mu_2} \dots \xi_{i_k}^{\mu_k} \xi_{j_k}^{\mu_k} \rangle \tag{A1.5}$$

with the restrictions $i_1 \neq j_1, \dots, i_k \neq j_k$. The non-vanishing contributions to this average are the fully contracted terms. These can be represented diagramatically as closed loops. A loop of order n has n vertices labelled by site indices. Adjacent vertices have different site indices. The contribution of such a loop is (as $N \to \infty$)

$$O(N^n (C/N)^{n'} C^{-n/2}). \tag{A1.6}$$

The first factor, N^n, is the number of groups of n sites. The second factor is the probability that the bonds, which are necessary to connect a given group n sites into a closed loop, actually exist in the network, the number of such bonds is $n' = 1$ for $n = 2$, and $n' = n$ otherwise. The third factor comes from the normalisation factor in (A1.1). The kth moment has loops of length k and, for $k \geq 4$, also products of shorter loops. These products are represented by disconnected diagrams and are always cancelled by the additional terms in the kth cumulant. This, is another example of the linked-cluster theorem which appears so often in statistical mechanics and many-body physics.

Thus, the contribution of the kth cumulant is of order N for $k = 2$, and of order $C^{k/2}$ for $k > 2$. Therefore, if C is of order N, as in a strongly connected network, the cumulant expansion does not converge. If, however, $C \leq \ln N$, all the terms $k > 2$ can be neglected as $N \to \infty$. This gives (10) of the text, in which only the $\langle z^2 \rangle$ term is retained.

Appendix 2

We calculate $\langle V^n \rangle$ for a replica- and site-symmetric solution, introducing the order parameters $q_{\alpha\beta}^i$ and $h_{\alpha\beta}^i$ by means of δ functions. Using the integral form of the δ functions, we can write, after rearrangement of terms,

$$\begin{aligned} \langle V^n \rangle = \frac{1}{S^n} \int \prod_{\alpha,i} (\mathrm{d}E_\alpha^i \, \mathrm{d}G_\alpha^i) \int \prod_{\alpha<\beta,i} \frac{\mathrm{d}F_{\alpha\beta}^i \, \mathrm{d}q_{\alpha\beta}^i}{2\pi} \int \prod_{\alpha\neq\beta,i} \frac{\mathrm{d}H_{\alpha\beta}^i \, \mathrm{d}h_{\alpha\beta}^i}{2\pi} \\ \times [\Theta(q_{\alpha\beta}^i, h_{\alpha\beta}^i)]^{\alpha C} \Phi(E_\alpha^i, G_\alpha^i, F_{\alpha\beta}^i, H_{\alpha\beta}^i) \exp\left(\tfrac{1}{2} C \sum_{\alpha,i} E_\alpha^i \right) \\ \times \exp\left(\tfrac{1}{2}\eta C \sum_{\alpha,i} G_\alpha^i - C \sum_{\alpha<\beta,i} F_{\alpha\beta}^i q_{\alpha\beta}^i - C \sum_{\alpha<\beta,i} H_{\alpha\beta}^i h_{\alpha\beta}^i \right) \end{aligned} \tag{A2.1}$$

where

$$\Theta(q^i_{\alpha\beta}, h^i_{\alpha\beta}) = \int_\kappa^\infty \prod_{\alpha,i} \mathrm{d}\lambda_i^\alpha \int_{-\infty}^\infty \prod_{\alpha,i} \frac{\mathrm{d}x_i^\alpha}{2\pi} \exp\left(\mathrm{i} \sum_{\alpha,i} x_i^\alpha \lambda_i^\alpha - \frac{1}{2} \sum_{\alpha,i} (x_i^\alpha)^2 \right)$$
$$\times \exp\left(-\frac{1}{2} \sum_{\alpha \neq \beta, i} q^i_{\alpha\beta} x_i^\alpha x_i^\beta - \frac{\eta}{2N} \sum_{\alpha,i,j} x_i^\alpha x_j^\alpha - \frac{1}{2N} \sum_{\alpha \neq \beta, i, j} h^i_{\alpha\beta} x_i^\alpha x_j^\beta \right) \quad \text{(A2.2)}$$

since the integrals over x and λ factorise over the patterns μ, where

$$\Phi(E^i_\alpha, G^i_\alpha, F^i_{\alpha\beta}, H^i_{\alpha\beta})$$
$$= \int \prod_{\alpha, i \neq j} \mathrm{d}J_{ij}^\alpha \exp\left(-\frac{1}{2} \sum_{\alpha,i} E^i_\alpha \sum_{j \neq i} (J_{ij}^\alpha)^2 \right)$$
$$\times \exp\left(\sum_{\alpha < \beta, i} F^i_{\alpha\beta} \sum_{j \neq i} J_{ij}^\alpha J_{ij}^\beta - \frac{1}{2} \sum_{\alpha,i} G^i_\alpha \sum_{j \neq i} J_{ij}^\alpha J_{ji}^\alpha + \frac{1}{2} \sum_{\alpha \neq \beta, i} H^i_{\alpha\beta} \sum_{j \neq i} J_{ij}^\alpha J_{ji}^\beta \right) \quad \text{(A2.3)}$$

and S is the denominator of (7), evaluated easily to give

$$S = \exp(CN/2). \quad \text{(A2.4)}$$

For replica- and site-symmetric solutions, i.e. for all i, α and $\alpha \neq \beta$,

$$\begin{aligned} E^i_\alpha &= E \qquad & q^i_{\alpha\beta} &= q \qquad & F^i_{\alpha\beta} &= F \\ G^i_\alpha &= G \qquad & h^i_{\alpha\beta} &= h \qquad & H^i_{\alpha\beta} &= H \end{aligned} \quad \text{(A2.5)}$$

we can express the integrand of (A2.1) as

$$\exp(\alpha C G_1(q, h) + \tfrac{1}{2} CN G_2(E, F, G, H))$$
$$\times \exp CN \left(\frac{n}{2} E - \frac{n(n-1)}{2} qF + \frac{n}{2} \eta G - \frac{n(n-1)}{2} hH \right) \quad \text{(A2.6)}$$

where

$$G_1(q, h) = \ln \int_\kappa^\infty \prod_{\alpha,i} \mathrm{d}\lambda_i^\alpha \int_{-\infty}^\infty \prod_{\alpha,i} \frac{\mathrm{d}x_i^\alpha}{2\pi} \exp\left(\mathrm{i} \sum_{\alpha,i} x_i^\alpha \lambda_i^\alpha - \frac{1}{2} \sum_{\alpha,i} (x_i^\alpha)^2 - \frac{q}{2} \sum_{\alpha \neq \beta, i} x_i^\alpha x_i^\beta \right)$$
$$\times \exp\left(-\frac{\eta}{2N} \sum_{\alpha,i,j} x_i^\alpha x_j^\alpha - \frac{h}{2N} \sum_{\alpha \neq \beta, i, j} x_i^\alpha x_j^\beta \right) \quad \text{(A2.7)}$$

$$G_2(E, F, G, H)$$
$$= \ln \int \prod_\alpha [\mathrm{d}J_1^\alpha \, \mathrm{d}J_2^\alpha] \exp\left(-\tfrac{1}{2} E \sum_\alpha [(J_1^\alpha)^2 + (J_2^\alpha)^2] \right)$$
$$\times \exp\left(\tfrac{1}{2} F \sum_{\alpha \neq \beta} (J_1^\alpha J_1^\beta + J_2^\alpha J_2^\beta) - G \sum_\alpha J_1^\alpha J_2^\alpha + H \sum_{\alpha \neq \beta} J_1^\alpha J_2^\beta \right) \quad \text{(A2.8)}$$

since the integrals over J_{ij} factorise over pairs of sites (i, j). For $n \ll 1$ we find

$$G_1(q, h) = nN \left[-\tfrac{1}{2} t_0^2 + \int_{-\infty}^\infty \mathrm{D}t \ln H \left(\frac{\kappa + \sqrt{\eta - h}\, t_0 + \sqrt{q}\, t}{\sqrt{1-q}} \right) \right] \quad \text{(A2.9)}$$

where t_0 is determined by a saddle-point equation, and

$$G_2(E, F, G, H) = -\frac{n}{2}\ln(E+F+G+H)+\frac{n}{2}\frac{F+H}{E+F+G+H} -\frac{n}{2}\ln(E+F-G-H)+\frac{n}{2}\frac{F-H}{E+F-G-H}. \tag{A2.10}$$

Setting E, F, G and H at their saddle-point values, we finally have

$$\langle V^n\rangle = \exp nCN\left(\tfrac{1}{4}\ln[(1-q)^2-(\eta-h)^2]+\frac{1}{2}\frac{q(1-q)-h(\eta-h)}{(1-q)^2-(\eta-h)^2}\right) \times \exp n\alpha CN\left[\int \mathrm{d}t \ln H\left(\frac{\kappa-r+\sqrt{q}\,t}{\sqrt{1-q}}\right)-\frac{1}{2}\frac{r^2}{\eta-h}\right] \tag{A2.11}$$

having defined

$$r = -\sqrt{\eta-h}\,t_0. \tag{A2.12}$$

The typical fractional volume is found by evaluating q, h and r at their respective saddle points and using

$$\langle \ln V\rangle = \lim_{n\to 0}\frac{\langle V^n\rangle - 1}{n} \tag{A2.13}$$

to obtain (13).

References

Abbott L F and Kepler T B 1988 Optimal learning in neural network memories *Preprint* Brandeis University BRX-TH-255

Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30

Crisanti A and Sompolinsky H 1988 *Phys. Rev.* A **37** 4865

Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257

Gardner E, Derrida B and Zippelius A 1987 *Europhys. Lett.* **4** 167

Gutfreund H, Reger J D and Young A P 1988 *J. Phys. A: Math. Gen.* **21** 2775

Hertz J A, Grinstein G and Solla S A 1987 *Heidelberg Coll. on Glassy Dynamics* ed J L van Hemmen and I Morgenstern (Berlin: Springer) p 538

Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554

Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380

Kohonen T 1984 *Self Organization and Associative Memory* (Berlin: Springer)

Krauth W, Nadal J-P and Mézard M 1988 *J. Phys. A: Math. Gen.* **21** 2995

Minsky M L and Papert S 1969 *Perceptrons* (Cambridge, MA: MIT Press)

Opper M 1988 Learning times of neural networks: exact solution for a perceptron algorithm *Preprint* University of Giessen, Julich

Parisi G 1986 *J. Phys. A: Math. Gen.* **19** L675

Personnaz L, Guyon I and Dreyfus G 1985 *J. Physique Lett.* **46** L359

Venkatesh S 1986 *PhD thesis* California Institute of Technology