J. Phys. A: Math. Gen. 22 (1989) 2181-2190. Printed in the UK

The space of interactions in neural networks: Gardner's computation with the cavity method

Marc Mézard

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

Received 16 January 1989

Abstract. Gardner's computation of the number of N-bit patterns which can be stored in an optimal neural network used as an associative memory is derived without replicas, using the cavity method. This allows for a unified presentation whatever the basic measure in the space of coupling constants, but above all it gives the clear physical content of the assumption of replica symmetry. TAP equations are also derived.

Foreword

One of the most exciting recent developments in the theory of neural networks is a contribution of Elizabeth Gardner. She showed how one can analyse the space of all the networks that are able to memorise a certain number of patterns, using spin-glass techniques and ideas. This was a major step forward and this kind of computation 'à la Gardner' (as it is often referred to among specialists) has been one of the most useful tools in the theory of neural networks since. As a tribute to my greatly missed colleague Elizabeth, to her talent and modesty, I decided to write down the following notes, which present an alternative and complementary derivation of her results.

1. Introduction

We consider a neural network of N binary neurons $\sigma_i = \pm 1$, $i = 1, \ldots, N$, fully connected (i.e. each neuron can interact with all the others). The dynamics of the network is deterministic:

$$\sigma_i^{t+1} = \operatorname{sgn}\left(\sum_j \tilde{J}_{ij}\sigma_j^t\right) \tag{1}$$

where $\tilde{J}_{ij}$ is the interaction from neuron j to neuron i, which need not be symmetric, and the updating can be either parallel or sequential, the type of dynamics having no influence on the issues which we shall study hereafter.

The network is used as an associative memory (Hopfield 1982) to store p patterns $\boldsymbol{\xi}^\mu = \{\xi_i^\mu = \pm 1,\ i = 1, \ldots, N\}$, $\mu = 1, \ldots, p$. Each network is characterised by the set of couplings $\{\tilde{J}_{ij}\}$. As two networks, which differ by an overall dilation of the couplings

† Laboratoire Propre du Centre National de la Recherche Scientique, associé à l'Ecole Normale Supérieure et à l'Université de Paris-Sud.

0305-4470/89/122181+10$02.50

($\tilde{J}_{ij} \to \lambda \tilde{J}_{ij}$), are clearly identical, we shall always suppose that the couplings have been normalised in such a way that the typical value of each $\tilde{J}_{ij}$ is of order $1 = N^0$ for $N \to \infty$. For instance we can study cases where $\tilde{J}_{ij} = \pm 1$ (Ising model), or $\tilde{J}_{ij}$ is real with $\Sigma_j \tilde{J}_{ij}^2 = N$ (spherical model), or $\tilde{J}_{ij} \in [-1, 1]$ (cubic model), etc. A given network stabilises the pattern μ if and only if $\boldsymbol{\xi}^\mu$ is a fixed point of the dynamics (1):

$$\xi_i^\mu = \operatorname{sgn}\left(\sum_j \tilde{J}_{ij} \xi_j^\mu\right) \qquad \forall i. \tag{2}$$

An important problem is the capacity of the network. Following the usual benchmarks (Hopfield 1982, Amit *et al* 1985), one can ask for instance how many random patterns ($\xi_i^\mu = \pm 1$ with probability $\frac{1}{2}$) can be stored by such a network. The answer obviously depends on the $\tilde{J}_{ij}$, and an important problem is to find good learning algorithms which provide a set of $\tilde{J}_{ij}$ that is able to memorise the largest possible number of patterns. But apart from any specific algorithm, it was recently realised by Gardner (1987, 1988) that one can compute the capacity of the best possible network.

The idea of her computation is, in the space of all the possible networks, to evaluate the fractional volume of the networks which stabilise all the patterns *on one given site* i_0. Obviously the stability equations (2) decouple on different sites, if no constraints (e.g. of symmetry) are placed *a priori* on the matrix of couplings. Denoting $\eta_j^\mu = \xi_{i_0}^\mu \xi_j^\mu$ and $J_j = \tilde{J}_{i_0 j}$, the stability equations of the p patterns on site i_0 reduce to

$$\Delta_\mu \equiv \frac{1}{\sqrt{N}} \sum_j J_j \eta_j^\mu > 0 \tag{3}$$

where we have introduced the stability parameters Δ_μ (Amari 1971, Gardner 1988, Krauth and Mézard 1987). The Δ_μ should be positive in order for the patterns to be stable, but it has also been shown that the larger the Δ_μ (in the spherical normalisation where $\Sigma_j J_j^2 = N$), the larger is the basin of attraction in the associative recall of pattern μ (Forrest 1988, Krauth *et al* 1988a, b, Kepler and Abbott 1988).

Hereafter we shall follow the approach of Gardner and Derrida (1988) where instead of demanding the full stability of the patterns, one introduces an energy

$$E\{J, \boldsymbol{\xi}^\mu\} = \sum_\mu \theta\left(K - \frac{1}{\sqrt{N}} \sum_j J_j \xi_j^\mu\right) \tag{4}$$

equal to the number of patterns with stability less than a given number K. Given a set of quenched random patterns $\boldsymbol{\xi}^\mu$, one seeks the set of couplings J_j which minimises the energy. The corresponding partition function at temperature $T = 1/\beta$ is:

$$\begin{aligned} Z_{N,p}\{\boldsymbol{\xi}^\mu\} &= \int \prod_j \rho(J_j)\, \mathrm{d}J_j \exp(-\beta E\{J, \boldsymbol{\xi}^\mu\}) \\ &= \int \prod_{j=1}^N \rho(J_j)\, \mathrm{d}J_j \prod_{\mu=1}^p \left[\mathrm{e}^{-\beta} + (1 - \mathrm{e}^{-\beta})\theta\left(\frac{1}{\sqrt{N}} \sum_j J_j \xi_j^\mu - K\right)\right] \end{aligned} \tag{5}$$

where we have written the natural measure on the J as $\Pi_j \rho(J_j)\, \mathrm{d}J_j$ (e.g. Ising measure: $\rho(J) = \frac{1}{2}[\delta(J+1) + \delta(J-1)]$). This does not exactly include the spherical model but it would be easy to incorporate this case into our framework by the use of an appropriate Lagrange multiplier.

This Gardner–Derrida approach allows us to analyse the space of couplings J_j: we have a kind of spin-glass Hamiltonian (4), where the basic thermalised variables are the couplings J_j, while the quenched parameters are the patterns ξ^μ, which give rise

to quenched frustrated random interactions between the J_j. Therefore we can compute the typical number of networks (related to the entropy) which produce a given number of errors (related to the energy), as well as more detailed pieces of information on the distribution of overlaps between the networks which stabilise a given set of patterns, etc. It should be emphasised that this approach can give much more detailed results than just the capacity. For the spherical model $\Sigma_j J_j^2 = N$ the stabilising volume is convex and the capacity is equal to $2N$ (Gardner 1987, 1988) for $N \to \infty$. This is a rather simple case for which good algorithms can reach the limit of optimality (Rosenblatt 1962, Minsky and Pappert 1969, Gardner *et al* 1987, Diedrich and Opper 1987, Poppel and Krey 1987) and even produce the largest possible stabilities (Krauth and Mézard 1987). But for cases that are of great practical importance, like the Ising case ($J_j = \pm 1$), the situation is much more complicated, as much from the analytic side (Gardner and Derrida 1988) as from the numerical side (Amaldi and Nicolis 1988): the capacity is still unknown, not to speak of the structure of the stabilising volume (overlap distribution, etc).

Hereafter we shall use the cavity method (Mézard *et al* 1987) to study the thermodynamics of this problem. We shall essentially work within the approximation of one single pure state (replica symmetric approximation) and point out the physical meaning of this approximation. There are two equivalent ways to use the cavity method on this problem. One way would be to use integral representations of the θ functions in (5) like Gardner (1987):

$$\theta\left(\frac{1}{\sqrt{N}}\sum_i J_i\xi_i^\mu - K\right) = \int_{-\infty}^{\infty}\frac{\mathrm{d}\lambda^\mu}{2\pi}\int_K^\infty \mathrm{d}t^\mu \exp\left[\mathrm{i}\lambda^\mu\left(t^\mu - \frac{1}{\sqrt{N}}\sum_i J_i\xi_i^\mu\right)\right] \tag{6}$$

and consider the problem of the $N+2p$ variables J_i, λ^μ, t^μ interacting through some couplings ξ_i^μ. This is probably the easiest way but it hides the physical content of the method. Therefore herafter we shall avoid this manipulation and work only in terms of the original 'physical' variables J_i. The cavity method will be implemented in two steps: adding a new pattern, or adding a new coupling. We shall always work in the limit $N \gg 1$, with $\alpha = P/N$ fixed. Eventually in the final section we shall derive TAP equations.

2. The cavity method: adding one constraint

To the problem with N couplings and p patterns (i.e. constraints) defined by the partition function (5) we add one new pattern $\xi_i^0 = \pm 1$, $i = 1, \ldots, N$. The new partition function is

$$Z_{N,p+1} = \int \prod_i \rho(J_i)\,\mathrm{d}J_i \prod_{\mu=1}^{p}\left[\mathrm{e}^{-\beta} + (1-\mathrm{e}^{-\beta})\theta\left(\frac{1}{\sqrt{N}}\sum_i J_i\xi_i^\mu - K\right)\right]$$
$$\times\left[\mathrm{e}^{-\beta} + (1-\mathrm{e}^{-\beta})\theta\left(\frac{1}{\sqrt{N}}\sum_i J_i\xi_i^0 - K\right)\right]. \tag{7}$$

The stability of the new pattern is:

$$\Delta_0 = \frac{1}{\sqrt{N}}\sum_i J_i\xi_i^0. \tag{8}$$

It has a certain thermal distribution due to the thermal fluctuations of the variables J_i. Denoting by $\langle\ \rangle$ the thermal averages (and, for future use, by $\overline{(\ )}$ the quenched averages over the distribution of random patterns), we get the first two moments:

$$h = (\Delta_0) = \frac{1}{\sqrt{N}} \sum_i \langle J_i \rangle \xi_i^0$$

$$\langle \Delta_0^2 \rangle - \langle \Delta_0 \rangle^2 = \frac{1}{N} \sum_i (\langle J_i^2 \rangle - \langle J_i \rangle^2) + \frac{1}{N} \sum_{i \neq j} \xi_0^i \xi_0^j (\langle J_i J_j \rangle - \langle J_i \rangle \langle J_j \rangle). \tag{9}$$

Following Mézard *et al* (1988), within one pure state the connected correlation functions like $\langle J_i J_j \rangle - \langle J_i \rangle \langle J_j \rangle$ should be generically small (e.g. of order $1/\sqrt{N}$), a property known as clustering which is expressed more rigorously by

$$\lim_{N \to \infty} \frac{1}{N^2} \sum_{i \neq j} (\langle J_i J_j \rangle - \langle J_i \rangle \langle J_j \rangle)^2 = 0. \tag{10}$$

Restricting the thermal measure to within one pure state, such that this clustering holds (as well as similar relations on higher-order correlations), one finds that the last term in (9) is negligible (since ξ_0^i and ξ_o^j are totally uncorrelated with $\langle J_i J_j \rangle - \langle J_i \rangle \langle J_j \rangle$), and the higher moments are also easily deduced: the thermal distribution of Δ_0 turns out to be a Gaussian, truncated by the new constraint introduced in (7):

$$P_{N,p+1}(\Delta_0) = \frac{c'}{\sqrt{2\pi(q_1 - q_0)}} \exp\left(-\frac{(\Delta_0 - h)^2}{2(q_1 - q_0)} \right) [e^{-\beta} + (1 - e^{-\beta}) \theta(\Delta_0 - K)] \tag{11}$$

where

$$q_1 = \frac{1}{N} \sum_i \langle J_i^2 \rangle = \overline{\langle J_i^2 \rangle}$$

$$q_0 = \frac{1}{N} \sum_i \langle J_i \rangle^2 = \overline{\langle J_i \rangle^2}. \tag{12}$$

The average h of Δ_0, defined in (9), depends on the pure state and on the sample. To proceed with relatively simple formulae, let us assume that there is only one pure state, in which case $\langle\ \rangle$ is the full Gibbs measure and h fluctuates only from sample to sample. The distribution of h is easily deduced from

$$\bar{h} = 0 \qquad \overline{h^2} = \frac{1}{N} \sum_i \overline{\langle J_i \rangle^2} = q_0 \qquad \text{etc.} \tag{13}$$

It is a Gaussian of mean 0 and width $\sqrt{q_0}$. Hereafter we shall denote this measure $D_{q_0}(h)$:

$$D_{q_0}(h) \equiv \frac{dh}{\sqrt{2\pi q_0}} \exp\left(-\frac{h^2}{2q_0} \right). \tag{14}$$

Let us suppose for a while that we know q_0 and q_1. Then the internal energy is:

$$E = \overline{\sum_\mu \left\langle \theta\left(K - \sum_i \frac{J_i \xi_i^\mu}{\sqrt{N}} \right) \right\rangle} = \alpha N \overline{\langle \theta(K - \Delta_0) \rangle}$$

$$= \alpha N \int D_{q_0}(h) \int D_{q_1 - q_0}(\Delta_0 - h)\, e^{-\beta} \theta(K - \Delta_0)$$

$$\times \left(e^{-\beta}(1 - e^{-\beta}) \int D_{q_1 - q_0}(\Delta_0 - h) \theta(\Delta_0 - K) \right)^{-1}. \tag{15}$$

It is useful to introduce the error function:

$$H(x) = \int_x^\infty \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{16}$$

in terms of which the energy is

$$E = \alpha N \int D_{q_0}(h)\, e^{-\beta}\left[1 - H\left(\frac{K-h}{\sqrt{q_1 - q_0}}\right)\right]\left[e^{-\beta} + (1-e^{-\beta})H\left(\frac{K-h}{\sqrt{q_1-q_0}}\right)\right]^{-1}. \tag{17}$$

We are now left with the computation of q_0 and q_1 which will be done self-consistently in the next section by adding one coupling instead of one constraint.

3. The cavity method: adding one coupling

Starting from the system with N couplings and p patterns, defined in (5), we add one new coupling J_0. The values of the patterns on the new site are $\xi_{i=0}^\mu = \pm 1$, $\mu = 1, \ldots, p$. The stability of each pattern is slightly changed, by a term of order $1/\sqrt{N}$. This suggests the expansion

$$\theta\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N J_i\xi_i^\mu - K + \frac{1}{\sqrt{N}} J_0\xi_0^\mu\right) = \theta\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N J_i\xi_i^\mu - K\right) + \varepsilon^\mu \tag{18}$$

where

$$\varepsilon^\mu = \frac{J_0\xi_0^\mu}{\sqrt{N}}\,\delta\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N J_i\xi_i^\mu - K\right) + \frac{1}{2!}\left(\frac{J_0\xi_0^\mu}{\sqrt{N}}\right)^2 \delta'\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N J_i\xi_i^\mu - K\right) + \ldots. \tag{19}$$

Using (5) and (18), we obtain the thermal distribution of the new coupling J_0, in the system with $N+1$ couplings and p patterns:

$$\mathcal{P}_{N+1,p}(J_0) = c'\rho(J_0)\int \prod_j \rho(J_j)\,dJ_j \sum_{k=0}^p \sum_{1\leqslant\nu_1\leqslant<\ldots<\nu_k\leqslant p} \varepsilon_{\nu_1}\ldots\varepsilon_{\nu_k}(1-e^{-\beta})^k$$
$$\times \prod_{\mu(\neq\nu_1,\ldots,\nu_k)}\left[e^{-\beta} + (1-e^{-\beta})\theta\left(\frac{1}{\sqrt{N}}\sum_i J_i\xi_i^\mu - K\right)\right]. \tag{20}$$

The information we need on the system with N couplings concerns the joint probability distribution of the stabilities

$$P_{N,p}^{\nu_1\ldots\nu_k}(\Delta_1,\ldots,\Delta_k) = \frac{1}{Z_{N,p}}\int \prod_j \rho(J_j)\,dJ_j \prod_{a=1}^k \delta\left(\Delta_{\nu_a} - \frac{1}{\sqrt{N}}\sum_i J_i\xi_i^{\nu_a}\right)$$
$$\times \prod_{\mu(\neq\nu_1,\ldots,\nu_k)}\left[e^{-\beta} + (1-e^{-\beta})\theta\left(\frac{1}{\sqrt{N}}\sum_i J_i\xi_i^\mu - K\right)\right]. \tag{21}$$

The clustering property means that inside one pure state this probability generically factorises into the product of individual distributions of each stability (by generically we mean that clustering holds on average for all moments, as in (10):

$$\lim_{N\to\infty}\frac{1}{p^2}\sum_{\mu\neq\nu}(\langle\Delta_\mu\Delta_\nu\rangle - \langle\Delta_\mu\rangle\langle\Delta_\nu\rangle)^2 = 0 \tag{22}$$

as well as generalisations of this formula to higher moments).

Then, the thermal distribution of J_0 inside one pure state is:

$$\mathscr{P}_{N+1,p}(J_0) = c'\rho(J_0) \prod_{\mu=1}^{p} \left\{ \int P^{\mu}_{N,p}(\Delta_\mu)\, d\Delta_\mu \left[1 + (1-e^{-\beta}) \frac{J_0 \xi_0^\mu}{\sqrt{N}} \delta(\Delta_\mu - K) + \frac{(1-e^{-\beta})}{2!} \left(\frac{J_0 \xi_0^\mu}{\sqrt{N}} \right)^2 \delta'(\Delta_\mu - K) + \ldots \right] \right\}. \tag{23}$$

In the large-N limit the higher-order terms in $1/\sqrt{N}$ become irrelevant and we get

$$\mathscr{P}_{N+1,p}(J_0) = c'\rho(J_0) \exp\left((1-e^{-\beta}) J_0 \sum_\mu \frac{\xi_0^\mu}{\sqrt{N}} P^{\mu}_{N,p}(K) \right) \times \exp\{\tfrac{1}{2}\alpha J_0^2 [(1-e^{-\beta}) \overline{P'^{\mu}_{N,p}(K)} - (1-e^{-\beta})^2 \overline{(P^{\mu}_{N,p}(K))^2}]\}. \tag{24}$$

Physically the situation is rather clear. By adding a new coupling we slightly change the stability Δ_μ of each pattern. If the Δ_μ (in the system with N couplings) is far from the boundary value K, then nothing changes. The only non-trivial constraints on J_0 come from those patterns for which Δ_μ was nearly equal to K. Therefore the distribution of J_0 depends on the probability distribution of stabilities around the value K.

But this probability distribution we know from the first part of the cavity method—the one in which we added one new pattern: it is explicitly written in (11) and (14). Therefore the relevant parameters are easily computed:

$$\begin{aligned} A &\equiv \alpha(1-e^{-\beta})^2 \overline{[P^{\mu}_{N,p}(K)]^2} \\ &= \alpha(1-e^{-\beta})^2 [P^{\mu=0}_{N,p+1}(K)]^2 \\ &= \alpha \int D_{q_0}(h) \left\{ \frac{1-e^{-\beta}}{\sqrt{2\pi(q_1-q_0)}} \exp\left(-\frac{(K-h)^2}{2(q_1-q_0)} \right) \times \left[e^{-\beta} + (1-e^{-\beta}) H\left(\frac{K-h}{\sqrt{q_1-q_0}} \right) \right]^{-1} \right\}^2 \end{aligned} \tag{25}$$

$$\begin{aligned} B &\equiv \alpha(1-e^{-\beta}) \overline{P'^{\mu=0}_{N,p+1}(K)} \\ &= \alpha \int D_{q_0}(h) \frac{K-h}{q_1-q_0} (1-e^{-\beta}) \frac{1}{\sqrt{2\pi(q_1-q_0)}} \\ &\quad \times \exp\left(-\frac{(K-h)^2}{2(q_1-q_0)} \right) \left[e^{-\beta} + (1-e^{-\beta}) H\left(\frac{K-h}{\sqrt{q_1-q_0}} \right) \right]^{-1}. \end{aligned} \tag{26}$$

Apart from these parameters A and B, the distribution of J_0 in (24) also depends on a field:

$$H = (1-e^{-\beta}) \sum_\mu \frac{\xi_0^\mu}{\sqrt{N}} P^{\mu}_{N,p}(K). \tag{27}$$

If there is only one pure state, this field is only sample dependent, and one gets $\bar{H} = 0$, $\overline{H^2} = A$, etc, so that H is a Gaussian variable of mean 0 and width $\sqrt{A}$.

Now we completely know the distribution (24) of J_0, from which we can deduce the values of q_0 and q_1, in (12):

$$q_0 = \overline{\langle J_0 \rangle^2} = \int D_A(H) \left(\frac{\int \rho(J_0)\, dJ_0 \exp[J_0 H + \frac{1}{2}(B-A)J_0^2] J_0}{\int \rho(J_0)\, dJ_0 \exp[J_0 H + \frac{1}{2}(B-A)J_0^2]} \right)^2 \tag{28}$$

$$q_1 = \overline{\langle J_0^2 \rangle} = \int D_A(H) \left(\frac{\int \rho(J_0)\, dJ_0 \exp(J_0 H + \frac{1}{2}(B-A)J_0^2] J_0^2}{\int \rho(J_0)\, dJ_0 \exp[J_0 H + \frac{1}{2}(B-A)J_0^2]} \right). \tag{29}$$

The set of four equations (25), (26), (28) and (29) defines the four parameters A, B, q_0 and q_1. Once these equations are solved, the whole thermodynamics can be recovered from expression (15) of the external energy as a function of temperature.

4. TAP equations

The previous results were derived by averaging over many samples, within the assumption of the existence of one single pure state. However, as explained by Mézard *et al* (1987) the cavity method can be used to study one single sample (always in the thermodynamic limit). In the present section we shall derive TAP-like equations (Thouless *et al* 1977) which are mean-field equations valid for one given sample, within one pure state.

Considering a sample with N couplings J_i and p patterns $\{\xi_i^\mu\}$, we shall need the following order parameters:

$$m_j = \langle J_j \rangle_{N,p} \qquad x^\mu_{N,p} = (1-e^{-\beta}) P^\mu_{N,p}(K) \tag{30}$$

where the thermal average is restricted to within one given pure state.

In § 3 we considered the addition of one new coupling J_0 to a system with N couplings and p constraints. We derived the distribution (24) of J_0 within one pure state (assumption of clustering). From (24) we have

$$\langle J_0^k \rangle_{N+1,p} = f_k(H_{N,p}) \tag{31}$$

where

$$f_k(H) = \frac{\int \rho(J_0)\, dJ_0 \exp[J_0 h + \frac{1}{2}(B-A)J_0^2] J_0^k}{\int \rho(J_0)\, dJ_0 \exp[J_0 h + \frac{1}{2}(B-A)J_0^2]} \tag{32}$$

and the field $H_{N,p}$ has been expressed in (17):

$$H_{N,p} = \sum_{\mu=1}^{p} \frac{\xi_0^\mu}{\sqrt{N}} x^\mu_{N,p}. \tag{33}$$

Equations (31) are cavity equations: they express the distribution of the new couplings J_0, in a sample with $N+1$ couplings, as a function of the x^μ in a sample of N couplings only. Therefore this set of equations is not closed. One needs to express $\langle J_0 \rangle_{N+1,p}$ in terms of the variables $x^\mu_{N+1,p}$ computed in the same sample. Let us express the probability $P^\mu_{N+1,p}(K)$ that the stability of pattern μ be equal to K in the system with $N+1$ couplings:

$$P^\mu_{N+1,p}(K) = \frac{1}{Z_{N+1,p}} \int \prod_{j=0}^{N} \rho(J_j)\, dJ_j\, \delta\left(K - \frac{1}{\sqrt{N}} \sum_j J_i \xi_j^\mu - \frac{1}{\sqrt{N}} J_0 \xi_0^\mu\right)$$
$$\times \prod_{\nu(\neq\mu)} \left[e^{-\beta} + (1-e^{-\beta}) \theta\left(\frac{1}{\sqrt{N}} \sum_{j=1}^{N} J_j \xi_j^\mu + \frac{1}{\sqrt{N}} J_0 \xi_0^\mu - K \right) \right]. \tag{34}$$

Following the approach of § 3, we expand all the terms in powers of $J_0 \xi_0^\mu / \sqrt{N}$. Introducing the joint probability distributions of the stabilities (21) and using the clustering hypothesis (22) we obtain eventually (we skip the details which are similar to the derivation of (23)):

$$P^\mu_{N+1,p}(K) = P^\mu_{N,p}(K) + \frac{1}{\sqrt{N}} \xi_0^\mu \langle J_0 \rangle_{N+1,p} [P'^\mu_{N,p}(K) - (1-e^{-\beta})(P^\mu_{N,p}(K))^2] \tag{35}$$

from which we obtain

$$H_{N+1,p} = \sum_{\mu=1}^{p} \frac{\xi_0^\mu}{\sqrt{N}} x^\mu_{N+1,p} = H_{N,p} + \langle J_0 \rangle_{N+1,p}(B-A). \tag{36}$$

Equations (31) and (36) give the first set of TAP equations; they relate the parameters m_j and x_μ in the same sample by

$$m_j \equiv \langle J_j \rangle = f_1\left(\sum_{\mu=1}^{p} \frac{\xi_j^\mu}{\sqrt{N}} x^\mu - (B-A)m_j \right). \tag{37}$$

The last term in the argument of f_1 is Onsager's reaction term.

We need a second set of equations expressing x^μ in terms of m_j. Adding one new pattern to a problem with N couplings and p patterns, we derived in § 2 the distribution of the stability of the new pattern. The analogue equation (31) gives the kth derivative of the probability that the stability of the new pattern be equal to K as

$$(1-e^{-\beta}) P^{(k)\,\mu=0}_{N,p+1}(K) = g_k(h_{N,p}) \tag{38}$$

where

$$g_k(h) = \frac{1-e^{-\beta}}{\sqrt{2\pi(q_1-q_0)}} \left[\left(\frac{\partial}{\partial t} \right)^{k-1} \exp\left(-\frac{(t-h)^2}{2(q_1-q_0)} \right) \right]\Bigg|_{t=K} \times \left[e^{-\beta} + (1-e^{-\beta}) H\left(\frac{K-h}{\sqrt{q_1-q_0}} \right) \right]^{-1} \tag{39}$$

and

$$h_{N,p} = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \xi_j^0 \langle J_j \rangle_{N,p}. \tag{40}$$

As before, we need to express $x^{\mu=0}_{N,p+1}$, and therefore $P^{\mu=0}_{N,p+1}$, in terms of $\langle J_j \rangle_{N,p+1}$. We have

$$\begin{aligned} h_{N,p+1} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i^0 \langle J_i \rangle_{N,p+1} \\ &= \frac{1}{Z_{N,p+1}} \int \sum_{j=1}^{N} \rho(J_j)\, dJ_j \frac{1}{\sqrt{N}} \sum_i J_i \xi_i^0 \\ &\quad \times \prod_{\mu=0}^{p} \left[e^{-\beta} + (1-e^{-\beta}) \theta\left(\frac{1}{\sqrt{N}} \sum_i J_i \xi_i^\mu - K \right) \right]. \end{aligned} \tag{41}$$

Following the method of § 2, this is easily written as

$$h_{N,p+1} = \int P_{N,p+1}(\Delta_0) \Delta_0 \, d\Delta_0 \tag{42}$$

where $P_{N,p+1}(\Delta_0)$ is given in (11). We thus find

$$h_{N,p+1} = h_{N,p} + (q_1-q_0)(1-e^{-\beta}) P^{\mu=0}_{N,p+1}(K) \tag{43}$$

which gives the second set of TAP equations (written in closed form for a system of N couplings and p patterns):

$$x^\mu = g_1\left(\frac{1}{\sqrt{N}} \sum_j \xi_j^\mu m_j - (q_1-q_0) x^\mu \right). \tag{44}$$

This ends the derivation of TAP equations. At this stage it may be useful to summarise the results. In one given sample, the order parameters are the average value m_j of the coupling J_j and a quantity x^μ proportional to the probability that a constraint be strict (see (30)). These order parameters satisfy the equations:

$$\begin{aligned} m_i &= f_1\left(\frac{1}{\sqrt{N}}\sum_\mu \xi_i^\mu x_\mu - (B-A)m_i\right) \qquad \forall i = 1,\ldots,N \\ x^\mu &= g_1\left(\frac{1}{\sqrt{N}}\sum_i \xi_i^\mu m_i - (q_1-q_0)x_\mu\right) \qquad \forall \mu = 1,\ldots,p \end{aligned} \tag{45}$$

where the parameters q_0, q_1, A and B, are given by

$$\begin{aligned} q_0 &= \frac{1}{N}\sum_i m_i^2 \\ q_1 &= \frac{1}{N}\sum_i f_2\left(\frac{1}{\sqrt{N}}\sum_\mu \xi_i^\mu x^\mu - (B-A)m_i\right) \\ A &= \frac{1}{N}\sum_\mu (x^\mu)^2 \\ B &= \frac{1}{N}\sum_\mu g_2\left(\frac{1}{\sqrt{N}}\sum_i \xi_i^\mu m_i - (q_1-q_0)x^\mu\right) \end{aligned} \tag{46}$$

and the functions f_k and g_k are given in (32) and (39).

5. Conclusions

The results of §§ 2 and 3 can also be obtained with the replica method. They have been derived in this way, for the cases of the spherical model or the Ising model, by Gardner and Derrida (1988). I think that, once again, the replica method and the cavity method are complementary. The former is much more compact and may be more systematic, the latter puts more accent on the physical content. For instance, keeping to the replica symmetric approximation, we have seen here that this abstract mathematical hypothesis hides two hypotheses:

the fact that the couplings are weakly correlated (clustering in the space of couplings, see (10));

the fact that different constraints (related to different patterns) are weakly correlated (clustering in the space of the stabilities, see (22)).

Whether these hypotheses are satisfied or not for a given $\rho(J)$ is related to the validity of the replica symmetric approximation. A self-consistent check (equivalent to a local stability analysis in the replica method) can be done (see Mézard *et al* (1987) where a similar computation has been performed), but that goes beyond the aim of this paper.

The TAP equations described in equations (45) and (46) can be quite helpful in this respect: these equations are valid within one pure state. Whenever ergodicity breaking occurs (which must be the case, for instance, for Ising couplings $J_j = \pm 1$), this should be signalled by the appearance of several solutions of these equations. Therefore it will be useful to study the number of solutions of these equations either numerically or analytically.

I hope the present approach will prove helpful for the understanding of the difficult problems where $\rho(J)$ is discrete, like the long-standing problem of the Ising case.

Acknowledgments

It is a pleasure to acknowledge many stimulating discussions with Werner Krauth and Jonathan Yedidia.

References

Amaldi E and Nicolis S 1988 *Preprint* Rome University no 642
Amari S 1971, *Proc. IEEE* **59** 35
Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530
Diederich S and Opper M 1987 *Phys. Rev. Lett.* **58** 949
Forrest B M 1988 *J. Phys. A: Math. Gen.* **21** 245
Gardner E 1987 *Europhys. Lett.* **4** 481
—— 1988 *J. Phys. A: Math. Gen.* **21** 257
Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
Gardner E, Stroud N and Wallace D J 1987 *Preprint* Edinburgh University 87/394
Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
Kepler T B and Abbott L F 1988 *J. Physique* **49** 1657
Krauth W and Mézard M 1987 *J. Phys. A: Math. Gen.* **20** L745
Krauth W, Mézard M and Nadal J P 1988a *J. Phys. A: Math. Gen.* **21** 2995
—— 1988b *Complex Systems* **2** 387
Mézard M, Parisi G and Virasoro M A 1988 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
Minsky M and Papert S 1969 *Perceptrons* (Cambridge, MA: MIT Press)
Poppel G and Krey U 1987 *Europhys. Lett.* **4** 481
Rosenblatt F 1962 *Principles of Neurodynamics* (New York: Spartan)
Thouless D, Anderson P W and Palmer R G 1977 *Phil. Mag.* **35** 593