Replica symmetry breaking in neural networks with modified pseudo-inverse interactions

1991 J. Phys. A: Math. Gen. 24 5163

(http://iopscience.iop.org/0305-4470/24/21/026)

J. Phys. A: Math. Gen. 24 (1991) 5163–5180. Printed in the UK

Replica symmetry breaking in neural networks with modified pseudo-inverse interactions

V S Dotsenko† and B Tirozzi

Dipartimento di Matematica, Università di Roma 'La Sapienza', Piazzale Aldo Moro 2, 00185 Roma, Italy

Received 25 April 1991

Abstract. Replica symmetry breaking is studied in fully connected neural networks with modified pseudo-inverse interactions. The interaction matrix has an intermediate form between the Hebb learning rule and the pseudo-inverse one. At low temperatures there is a region of parameters where the replica-symmetric solution is stable while its entropy is negative. It indicates the existence of an alternative solution in which the replica symmetry is broken. The one-step replica symmetry-breaking solution is found and its properties are analysed.

1. Introduction

The phenomenon of replica symmetry breaking (RSB) discovered by Parisi (1979) in the mean field solution of the Sherrington and Kirkpatrick (SK) (1975) model of spin glasses is now sufficiently well studied and its physical interpretation is clear (see Mezard *et al* 1987). RSB has been proved to exist in other spin glass-like systems, including statistical models of neural networks. However, unlike spin glasses where RSB is the crucial characteristic of the low-temperature phase, it is generally believed that in neural networks is not very important.

A classical example is the model proposed by Hopfield (1982). It was shown (e.g. see Amit *et al* 1987) that although at a low enough temperature the replica-symmetric (RS) solution is unstable against RSB, that instability is very weak and the true RSB solution is not much different from the RS one. The phenomenon of RSB itself was proved for that model to be qualitatively similar to that in the SK model with a magnetic field below the line of instability of the RS solution, known as the Almeida and Thouless line (AT line) (Almeida and Thouless 1978).

Here we consider a model of neural networks in which the phenomenon of RSB appears to be of a qualitatively different kind.

The model under consideration is the fully connected neural network which, in a sense, is intermediate between the Hopfield model and the so-called pseudo-inverse model, studied by Personaz *et al* (1985) and by Kanter and Sompolinsky (1987) (section 2).

The motivation, the RS solution and the resulting phase diagram of the considered model have been studied in detail by Dotsenko *et al* (1991).

† On leave from the Landau Institute for Theoretical Physics, Academy of Sciences of the USSR, Moscow, USSR.

It can be shown (section 3) that there is a certain range of parameters where the RS solution is still stable, but its entropy is becoming negative (so that the zero-entropy line goes above the AT line). This indicates that in addition to the stable RS solution there exists another solution with presumably 'exotic' RSB, which 'takes' a part of the entropy. As a result, below the AT line a sort of 'first-order' phase transition could be expected (with a finite jump in the structure of the order parameter although without discontinuity of the free energy). The structure of the RSB state above the AT line could then be expected to be very different from the 'traditional' one in the SK model and in the Hopfield model, since it does not appear as a result of instability of the RS solution, but, in a sense, on its own, at a 'finite distance' from it.

A similar phenomenon has been already found in the Gardner and Derrida (1988) problem of maximal capacity in neural networks with Ising interactions by Krauth and Mezard (1989) (see also Gutfreund and Stein 1990).

Therefore, it would be reasonable to expect that this could be a rather general phenomenon for a certain class of spin glass-like systems, although up to now it has been unclear what kind of general property makes such a system exhibit this effect. An essential advantage of the present model is that the considered effect occurs in the region of the phase diagram where there are small parameters, and thus an analytic approach can be used.

In section 4 the appearance of the RSB solution is analysed in the 'normal' region of the phase diagram where the zero-entropy line goes below the AT line. It is shown that when one approaches the intersection of these lines the second-order phase transition in the RSB phase has a tendency to turn into a first-order one with a finite jump of the order parameter. It is also shown that the RSB solution has a tendency to become close to that of one-step RSB.

In section 5 the one-step RSB solution is found and it is shown that the region where it exists *does not* coincide with that defined by the zero-entropy line (when lowering the temperature the RSB solution appears *before* the entropy changes sign). Although the stability of the obtained RSB solution remains an open question, it is shown that its entropy is positive.

The results obtained are summarized in section 6.

2. The model

We consider the model which consists of N Ising spins σ_i $(i=1,\ldots,N)$ and is described by the Hamiltonian

$$H=\tfrac{1}{2}\sum_{ij} J_{ij}\sigma_i\sigma_j. \tag{1}$$

The interaction matrix is taken in the form

$$J_{ij}=\frac{1}{N}\sum_{\mu,\nu=1}^{p}\xi_i^{\mu}(\hat{1}+\lambda\hat{C})^{-1}_{\mu\nu}\xi_j^{\nu} \tag{2}$$

where

$$\hat{C}_{\mu\nu}=\frac{1}{N}\sum_i \xi_i^{\mu}\xi_i^{\nu} \tag{3}$$

and ξ_i^{μ} $(\mu=1,\ldots,p)$ are quenched uncorrelated patterns (so that the off-diagonal elements of the matrix $\hat{C}$, equation (3), are of the order $1/\sqrt{N}$). λ is the parameter of the model. If $\lambda=0$ the model (1), (2) turns into the Hopfield one, and at $\lambda\to\infty$ the structure of the interaction matrix (2) approaches that of the pseudo-inverse model studied by Personaz *et al* (1985) and Kanter and Sompolinsky (1987). The motivation for this particular choice of the J_{ij} (equation (2)) has been given by Dotsenko *et al* (1991). It could be obtained from the traditional Hebb learning rule via a local thermal noise iterative procedure, and in the RS solution it provides a substantial increase in capacity and quality of the retrieval.

The model will be studied in the thermodynamic limit where both $N\to\infty$ and $P\to\infty$ while the parameter $\alpha=P/N$ remains finite.

The free energy of the model is calculated in terms of the replica approach:

$$-\beta NF=\lim_{n\to 0}\frac{\langle\langle Z^n\rangle\rangle-1}{n} \tag{4}$$

where $\langle\langle\ldots\rangle\rangle$ denotes averaging over the random ξ_i^μ and Z^n is the replica partition function:

$$Z^n=\sum_{(\sigma_i^\rho)}\exp\left(-\frac{\beta}{2N}\sum_{\rho=1}^{n}\sum_{ij}^{N}\sum_{\mu\nu}^{P}\xi_i^\mu(\hat{1}+\lambda\hat{C})^{-1}_{\mu\nu}\xi_j^\nu\sigma_i^\rho\sigma_j^\rho\right). \tag{5}$$

Introducing the fields a_μ^ρ, Φ_i^ρ one obtains

$$Z^n=\int \mathrm{D}\boldsymbol{a}\int \mathrm{D}\boldsymbol{\Phi}\sum_{(\sigma_i^\rho)}\exp\left(-\frac{\beta}{2N}\sum_{\rho\mu}(a_\mu^\rho)^2-\frac{\beta\lambda}{2}\sum_{i\rho}(\Phi_i^\rho)^2+\beta\sum_{i\mu\rho}a_\mu^\rho\xi_i^\mu(\sigma_i^\rho+\mathrm{i}\lambda\Phi_i^\rho)\right) \tag{6}$$

where the following symbols have been introduced:

$$\mathrm{D}\boldsymbol{a}=\prod_{\mu\rho}\mathrm{d}a_\mu^\rho \qquad \mathrm{D}\boldsymbol{\Phi}=\prod_{i\rho}\mathrm{d}\Phi_i^\rho.$$

The term containing $\det(\hat{1}+\lambda\hat{C})$, which contribute an irrelevant constant, is omitted. The fields a_μ are connected with the usual overlaps

$$m_\mu=\frac{1}{N}\sum_i\xi_i^\mu\langle\sigma_i\rangle \tag{7}$$

as follows,

$$a_\mu=\frac{1}{N}\sum_\nu(\hat{1}+\lambda\hat{C})^{-1}_{\mu\nu}m_\nu \tag{8}$$

while the meaning of the site field is given by the relation

$$\Phi_i=\mathrm{i}\sum_\mu a_\mu\xi_i^\mu. \tag{9}$$

Following standard calculations similar to those of the Hopfield model (e.g. see Amit *et al* 1987), after averaging over the ξ_i^μ one obtains

$$\langle\langle Z^n\rangle\rangle=\int\prod\mathrm{d}a^\rho\int\mathrm{D}\boldsymbol{Q}\int\mathrm{D}\boldsymbol{R}\exp(-\beta Nnf(\boldsymbol{a},\boldsymbol{Q},\boldsymbol{R})) \tag{10}$$

with

$$\mathrm{D}\boldsymbol{Q}=\prod_{\rho,\gamma}\mathrm{d}Q_{\rho,\gamma} \qquad \mathrm{D}\boldsymbol{R}=\prod_{\rho,\gamma}\mathrm{d}R_{\rho,\gamma}$$

and where

$$
\begin{aligned}
-\beta Nnf(\boldsymbol{a}, \boldsymbol{Q}, \boldsymbol{R}) \\
&= -\frac{\beta N}{2}\sum_{\rho}(a^{\rho})^2 - \frac{\alpha N}{2}\operatorname{tr}\log(\hat{1}-\beta\hat{Q}) - \frac{\alpha\beta^2 N}{2}\sum_{\rho,\gamma} Q_{\rho,\gamma}R_{\rho,\gamma} \\
&\quad + N\log\left[\prod_{\rho}\left(\sum_{\sigma}\int \mathrm{d}\Phi^{\rho}\right)\exp\left(-\tfrac{1}{2}\beta\lambda\sum_{\rho}(\Phi^{\rho})^2 + \beta\sum_{\rho}a^{\rho}(\sigma^{\rho}+\mathrm{i}\lambda\Phi^{\rho})\right.\right. \\
&\quad \left.\left. + \tfrac{1}{2}\alpha\beta^2\sum_{\rho,\gamma}(\sigma^{\rho}+\mathrm{i}\lambda\Phi^{\rho})(\sigma^{\gamma}+\mathrm{i}\lambda\Phi^{\gamma})R_{\rho\gamma}\right)\right]. \qquad (11)
\end{aligned}
$$

When obtaining the above expression the $n\times n$ matrix $\hat{Q}$ has been defined as

$$
Q_{\rho\gamma} = \frac{1}{N}\sum_{i}(\sigma_i^{\rho}+\mathrm{i}\lambda\Phi_i^{\rho})(\sigma_i^{\gamma}+\mathrm{i}\lambda\Phi_i^{\gamma}) \qquad (12)
$$

and the matrix $R_{\rho\gamma}$ has been defined as a conjugate variable to equation (12).

In the above calculations it was assumed that the pattern with number 1 is expected to condense and therefore the parameter a^{ρ} in equation (11) has been defined as

$$
a^{\rho} \equiv a^{\rho}_{\mu=1}\xi^{\mu=1}.
$$

3. RS solution, entropy and stability

(*a*) The RS solution assumes that

$$
R_{\rho\gamma} = \begin{cases} R & \text{if } \rho\neq\gamma \\ R_0 & \text{if } \rho=\gamma \end{cases} \qquad (13a)
$$

$$
Q_{\rho\gamma} = \begin{cases} Q & \text{if } \rho\neq\gamma \\ Q_0 & \text{if } \rho=\gamma \end{cases} \qquad (13b)
$$

and $a^{\rho} = a$.

In the $n\to 0$ limit, after some algebra, one obtains from equation (11)

$$
\begin{aligned}
f(a, Q_0, Q, R, \Delta) \\
&= \tfrac{1}{2}\left(1+\frac{\lambda}{\Delta}\right)a^2 + \frac{\alpha}{2\beta}\ln[1-\beta(Q_0-Q)] - \frac{\alpha}{2}\frac{Q}{1-\beta(Q_0-Q)} + \frac{\alpha}{2}R\beta(Q_0-Q) \\
&\quad + \frac{\Delta-1}{2\lambda}Q_0 + \frac{\lambda\alpha R}{2\Delta} - \frac{\Delta-1}{2\lambda\Delta} + \frac{\ln\Delta}{2\beta} - \frac{1}{\beta}\overline{\ln\left(\cosh\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)\right)} \qquad (14)
\end{aligned}
$$

where we have defined

$$
\Delta = 1 + \lambda\alpha\beta(R_0-R) \qquad (15)
$$

and

$$
\overline{(\ldots)} \equiv \int \frac{\mathrm{d}z}{\sqrt{(2\pi)}}\exp(-z^2/2)(\ldots). \qquad (16)
$$

The resulting saddle-point equations for the parameters α, Q_0, Q, R and Δ are

$$C \equiv \beta(Q_0 - Q) = \frac{\beta}{\Delta^2}\overline{\cosh^{-2}\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)} - \frac{\lambda}{\Delta} \tag{17}$$

$$R = \frac{Q}{(1-C)^2} \tag{18}$$

$$\Delta = 1 + \frac{\alpha\lambda}{(1-C)} \tag{19}$$

$$a = \frac{1}{(\lambda+\Delta)}\overline{\tanh\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)} \tag{20}$$

$$Q_0\Delta^2 = 1 + \lambda^2\alpha R - \lambda(\lambda+2\Delta)a^2 - \frac{\lambda\Delta}{\beta} - 2\alpha\lambda R\frac{\beta}{\Delta}\overline{\cosh^{-2}\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)}. \tag{21}$$

The analysis of these equations and the resulting phase diagram in the space of parameters T, α and λ have been reported by Dotsenko *et al* (1991).

Here we will consider the system in the limit $T \ll 1$, $\alpha \ll 1$ and $\lambda \ll 1$ where equations (17)–(21) give $a \simeq 1$, $R \simeq 1$ and $Q_0 \simeq 1$.

(*b*) The entropy of the RS solution is easily obtained from equation (14):

$$S = -\frac{\partial f}{\partial T} = -\frac{\alpha}{2}\left(\frac{C}{1-C} + \ln(1-C)\right) - \tfrac{1}{2}\ln(\Delta) + \overline{\ln\left(\cosh\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)\right)}$$
$$-\overline{\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)\tanh\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)}. \tag{22}$$

In order to define when the entropy changes sign two limits should be considered:

(i) For $T \ll \alpha$ one gets from equation (17)

$$C \simeq \sqrt{\frac{2}{\pi\alpha}}\exp\left(-\frac{1}{2\alpha}\right) - \lambda \ll 1 \tag{23}$$

$$\Delta = 1 + \frac{\lambda\alpha}{1-C}. \tag{24}$$

For the entropy (equation (22)) one obtains

$$S \simeq -\frac{\alpha}{2}[C^2/2 + \lambda(1+C)] + T\frac{\pi^2}{6\sqrt{2\pi\alpha}}\exp\left(-\frac{1}{2\alpha}\right). \tag{25}$$

Then in the limit $\lambda \ll \exp(-1/2\alpha)$ (when the Hopfield model is recovered) the entropy becomes negative below

$$T_{S=0}(\alpha) \simeq \frac{6}{\pi^2}\sqrt{\alpha/2\pi}\exp\left(-\frac{1}{2\alpha}\right). \tag{26}$$

However, if λ is not too small

$$\exp\left(-\frac{1}{2\alpha}\right) \ll \lambda \ll 1 \tag{27}$$

and equation (25) indicates that the entropy is always negative. This means that for λ finite (equation (27)) the entropy changes sign at much higher temperatures.

(ii) $\alpha \ll T \ll 1$. In this case one gets from equation (22)

$$S = -\frac{\alpha}{2}\lambda + \frac{2}{T}\exp\left(-\frac{2}{T}\right). \tag{28}$$

Therefore, in this limit, the entropy becomes negative below

$$T_{S=0} \simeq \frac{T^*}{1+\frac{1}{2}T^* \ln(1/\alpha)} \tag{29}$$

where

$$T^* \simeq \frac{2}{\ln(4/\lambda)}. \tag{30}$$

In other words, for λ finite (equation (27)) the zero entropy line goes to zero logarithmically slow:

$$T_{S=0}(\alpha) \sim \frac{1}{\ln(1/\alpha)}. \tag{31}$$

(*c*) The stability analysis of the obtained RS solution is fully analogous to that of the Hopfield model (e.g. see Amit *et al* 1987). The line of instability of the RS solution can be shown to be given by the equation

$$\alpha\beta^2\overline{(\langle S^2\rangle - \langle S\rangle^2)} = (1-C)^2 \tag{32}$$

where $S = \sigma + \mathrm{i}\lambda\Phi$, and

$$\langle(\ldots)\rangle = \frac{\Sigma_\sigma \int \mathrm{d}\Phi(\ldots)\exp H(\sigma,\Phi)}{\Sigma_\sigma \int \mathrm{d}\Phi \exp H(\sigma,\Phi)}. \tag{33}$$

Here

$$H(\sigma,\Phi) = -\tfrac{1}{2}\lambda\beta\Delta(\Phi-\Phi_0(\sigma))^2 + \frac{\beta}{\Delta}(a+\sqrt{\alpha R}z) \tag{34}$$

and

$$\Phi_0(\sigma) = \frac{\mathrm{i}}{\Delta}\left(a+\sqrt{\alpha R}z + \frac{\Delta-1}{\lambda}\sigma\right). \tag{35}$$

From equation (32) one obtains that for $\alpha \to 0$ the line of instability of the RS solution (the AT line) *does not depend on* λ and is given by

$$T_{\mathrm{AT}} \simeq \sqrt{\frac{8\alpha}{9\pi}}\exp\left(-\frac{1}{2\alpha}\right) \tag{36}$$

which is the same as in the Hopfield model (Amit *et al* 1987). The qualitative behaviour of the lines $T_{S=0}$ and $T_{\mathrm{AT}}(\alpha)$ is summarized in figure 1. Therefore, in the region

$$\alpha < \alpha^* \simeq \frac{2}{\ln(1/\lambda)} \tag{37a}$$

$$T < T^* \simeq \frac{2}{\ln(1/\lambda)} \tag{37b}$$

the zero-entropy line $T_{S=0}(\alpha)$ is higher than the AT line and we are facing the situation that at $T_{\mathrm{AT}}(\alpha) < T < T_{S=0}(\alpha)$ the RS solution is stable while the entropy is negative.

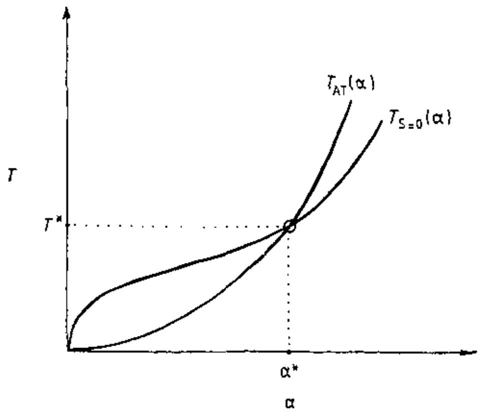

Figure 1.

4. 'Traditional' RSB near the AT line

In this section we investigate the RSB solution near the AT line in the region $\alpha > \alpha^*$ where the situation is 'normal', i.e. $T_{S=0}(\alpha) < T_{AT}(\alpha)$ (figure 1). Here one can expect that in the vicinity of the line T_{AT} the RSB is weak, and that the corrections to the RS solution are small. Therefore, in the exact expression for the free energy (11) we can put the replica matrix $\hat{R}$ in the form

$$R_{\rho\gamma} = \begin{cases} R + \dfrac{1}{\alpha\beta^2} r_{\rho\gamma} & \text{if } \rho \neq \gamma \\ R_0 & \text{if } \rho = \gamma \end{cases} \tag{38}$$

where the off-diagonal corrections are small $r_{\rho\gamma} \ll 1$. Then, according to the general scheme (e.g. see Parisi 1980) we should make an expansion over $\hat{r}$ up to fourth order.

In the $n \to 0$ limit, when the off-diagonal elements $Q_{\gamma\rho}$ and $r_{\rho\gamma}$ turn into the continuous functions $Q(x)$ and $r(x)$, after some algebra (see appendix 1) one obtains the following expression for the free energy:

$$\begin{aligned} f(a, Q_0, \Delta, Q(x), r(x)) & \\ = \frac{1}{2}\left(1+\frac{\lambda}{\Delta}\right) a^2 + \frac{\alpha\beta}{2} R Q_0 + \frac{\Delta-1}{2\lambda} Q_0 - \frac{\Delta-1}{2\lambda\Delta} + \frac{\ln\Delta}{2\lambda\Delta} + \frac{\lambda\alpha R}{2\Delta} & \\ - \frac{1}{\beta} \overline{\ln\left(\cosh\frac{\beta}{\Delta}(a+\sqrt{\alpha R} z)\right)} - \frac{\alpha}{2\beta}\int_n^1 \frac{\mathrm{d}x}{x^2} \ln(1-\chi(x)) & \\ + \frac{\alpha}{2\beta n} \ln(1-\chi(n)-\beta n Q(n)) - \frac{\alpha\beta}{2}\int_0^1 \mathrm{d}x\, Q(x)\left(R+\frac{1}{\alpha\beta^2} r(x)\right) & \\ + \frac{1}{2}\overline{\langle S\rangle^2}\int_0^1 \mathrm{d}x\, r(x) + \frac{1}{4}(a_0+2a_1)\int_0^1 \mathrm{d}x\, r(x)^2 & \\ - \frac{a_1}{2}\int_0^1 \mathrm{d}x\left(x r(x)^2 + \int_0^x \mathrm{d}x'\, r(x')^2 + 2r(x)\int_0^x \mathrm{d}x'\, r(x')\right) & \\ - \frac{b}{2}\int_0^1 \mathrm{d}x\left(x r(x)^3 + 3r(x)\int_0^x \mathrm{d}x'\, r(x')^2\right) & \\ + \frac{1}{6} c \int_0^1 \mathrm{d}x\, r(x)^3 + \frac{1}{8} d \int_0^1 \mathrm{d}x\, r(x)^4. & \end{aligned} \tag{39}$$

Only for third and fourth order are the terms responsible for the RSB kept. Here

$$\chi(x)=\beta\left(Q_0-xQ(x)-\int_x^1 \mathrm{d}x'\,Q(x')\right) \tag{40}$$

and the coefficients are as follows:

$$\begin{aligned} a_0&=\overline{\langle\!\langle S^2\rangle\!\rangle}\\ a_1&=\overline{\langle S\rangle^2\langle\!\langle S^2\rangle\!\rangle}\\ b&=\overline{\langle\!\langle S^2\rangle\!\rangle^3}\\ c&=\tfrac{1}{2}\overline{\langle\!\langle S^3\rangle\!\rangle^2}\\ d&=\tfrac{1}{6}\overline{\langle\!\langle S^4\rangle\!\rangle}. \end{aligned} \tag{41}$$

The notation $\langle\!\langle\ldots\rangle\!\rangle$ indicates the irreducible average according to equations (33)–(35), and $S=\sigma+\mathrm{i}\Phi$.

The calculations give

$$\begin{aligned} \overline{\langle S\rangle^2}&=\frac{1}{\Delta^2}\overline{\left(\tanh\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)-\lambda(a+\sqrt{\alpha R}z)\right)^2}\\ a_0&=\overline{\left(\frac{1}{\Delta^2}\cosh^{-2}\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)-\frac{\lambda}{\beta\Delta}\right)^2}\\ a_1&=\overline{\frac{1}{\Delta^2}\left(\tanh\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)-\lambda(a+\sqrt{\alpha R}z)\right)^2\left(\frac{1}{\Delta^2}\cosh^{-2}\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)-\frac{\lambda}{\beta\Delta}\right)}\\ b&=\overline{\left(\frac{1}{\Delta^2}\cosh^{-2}\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)-\frac{\lambda}{\beta\Delta}\right)^3}\\ c&=\frac{2}{\Delta^4}\overline{\tanh^2\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)\left(1-\frac{1}{\Delta}\tanh^2\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)\right)^2}. \end{aligned} \tag{42}$$

The coefficient $d>0$ and its explicit value is not very important for the results.

The variation of the free energy (39) gives the following saddle-point equations:

$$r(x)=\alpha\beta^2\int_{x_0}^{x}\mathrm{d}y\,\frac{q'(y)}{(1-\chi(y))^2} \tag{43}$$

$$R=\frac{q_0}{(1-\chi_0)^2} \tag{44}$$

$$\begin{aligned} Q(x)=\overline{\langle S\rangle^2}+a_0r(x)-2a_1\bar{r}-b\left(xr^2(x)+\int_0^x\mathrm{d}y\,r^2(y)+2r(x)\int_x^1\mathrm{d}y\,r(y)\right)\\ +cr^2(x)+dr^3(x) \end{aligned} \tag{45}$$

where $\bar{r}\equiv\int_0^1\mathrm{d}x\,r(x)$, and $\chi_0\equiv\chi(0)$.

The functions $Q(x)$ and $r(x)$ are assumed to be constants in the intervals $0\leqslant x\leqslant x_0$ and $x_1\leqslant x\leqslant 1$: $Q(x\leqslant x_0)=q_0$, $Q(x\geqslant x_1)=q_1$, $r(x\leqslant x_0)=0$ and $r(x\geqslant x_1)=r_1$ (figure 2).

Differentiating twice with respect to x one obtains that, in the interval where the functions are not constants,

$$3\,\mathrm{d}r(x)=\left(b+\frac{(1-\chi(x))^2}{\alpha\beta^3}\right)x-C. \tag{46}$$

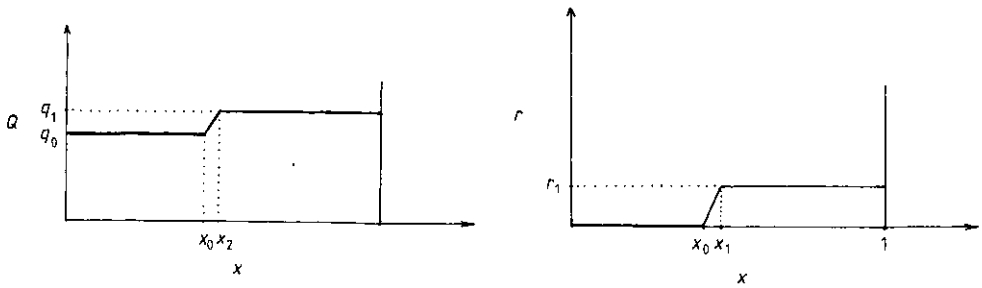

Figure 2.

Since $r(x)$ is assumed to be small we may restrict ourselves to a linear approximation on x which gives for $x_0 \leqslant x \leqslant x_1$

$$r(x) = A(x - x_0) \tag{47}$$

$$Q(x) = q_0 + \frac{(1-\chi_0)^2}{\alpha\beta^2} A(x - x_0)$$

where

$$A = \frac{1}{3d}\left(b + \frac{(1-\chi_0)^3}{\alpha\beta^3}\right) \tag{48}$$

$$x_0 = \frac{c}{[b + (1-\chi_0)^3/\alpha\beta^3]}. \tag{49}$$

For the intervals $0 \leqslant x \leqslant x_0$ and $x_1 \leqslant x \leqslant 1$, equation (45) gives

$$q_0 = \overline{\langle S\rangle^2} - 2a_1\bar{r} \tag{50}$$

$$q_1 = q_0 + a_0 r_1 - b(r_1^2 + \overline{r^2}) + c r_1^2 + d r_1^3$$

where, using equations (47) and (40),

$$\bar{r} \equiv \int_0^1 \mathrm{d}x\, r(x) = r_1(1 - x_0) + \frac{1}{2A} r_1^2$$

$$\overline{r^2} \equiv \int_0^1 \mathrm{d}x\, r(x)^2 = r_1^2(1 - x_0) + \frac{2}{3A} r_1^3$$

$$\chi_0 = \beta(Q_0 - \bar{q}) \tag{51}$$

$$\bar{q} \equiv \int_0^1 \mathrm{d}x\, Q(x) = q_0 + \frac{(1-\chi_0)^2}{\alpha\beta^2}\bar{r}$$

$$q_1 = q_0 + \frac{(1-\chi_0)^2}{\alpha\beta^2} r_1.$$

Inserting equations (51) into equations (50) one finally obtains

$$-\tau r_1 + \kappa r_1^2 - \gamma r_1^3 = 0 \tag{52}$$

where

$$\tau = a_0 - \frac{(1-\chi_0)^2}{\alpha\beta^2} \tag{53}$$

$$\kappa = b(2 - x_0) - C \tag{54}$$

$$\gamma = d\left(\frac{2b}{c} x_0 + 1\right). \tag{55}$$

Non-zero solutions for r_1 appear only for $\tau > 0$, and the equality $\tau = 0$ gives the AT line. For $T \ll 1$, $\alpha \ll 1$ and $\lambda \ll 1$, using equation (42) for a_0, one can then easily get the AT line explicitly,

$$T_{\mathrm{AT}} = \sqrt{\frac{8\alpha}{9\pi}} \exp\left(-\frac{1}{2\alpha}\right) \tag{56}$$

which to the main order does not depend on λ. At $\tau > 0$ ($T < T_{\mathrm{AT}}$) for $\tau \ll 1$

$$r_1 \simeq \frac{\tau}{\kappa} \tag{57}$$

and all the other parameters which describe the RSB solutions are defined in equations (48), (49) and (51). Qualitatively the obtained RSB solutions are shown in figure 2.

However, the solution for r_1 (equation (57)) makes sense only if $\kappa > 0$ and if the value of κ is not very small.

Using equations (42) for the parameters b and c one obtains the value of the parameters κ (equation (54)):

$$\kappa \simeq \exp\left(-\frac{1}{\alpha}\right)(2 - x_0) - \alpha^2\lambda^2 \tag{58}$$

where x_0 (equation (49)) is less than 1. The above equations show that when α approaches from above the value $\alpha^* \sim 1/\ln(1/\lambda)$, the coefficient $1/\kappa$ in equation (57) is diverging. At $\alpha < \alpha^*$ the parameter κ is negative, and the RSB solution of the form shown in figure 2 does not exist.

Actually, the fact that in the region $\alpha < \alpha^*$ the parameter $\kappa < 0$, means that the true RSB solution which appears below the AT line cannot be described as a *small* deviation from the RS solution. RSB should be expected to be already *finite* there. The form of this true solution is indicated by equations (47) and (48), which in a sense do not depend on the limiting value of r_1, and which shows how the function $r(x)$ behaves in the region where it is not constant. Equation (48) shows that at $\alpha \to 0$ the slope of the curve $r(x)$ is of the order $1/\alpha \to \infty$, i.e. the function $r(x)$ could be expected to be almost the vertical step. It means that $\alpha \to 0$ the true RSB solution should be expected to be close to that given by the one-step RSB.

5. One-step RSB

In this section we investigate the region $\alpha \ll \alpha^* \simeq \ln(1/\lambda)$ (figure 1). It will be shown that below the temperature $T^* \simeq 2/\ln(4/\lambda)$, after crossing the transition line $T_{\mathrm{RSB}}(\alpha) \simeq T^* - \alpha$ (when $\alpha \ll T^2$), there exists a one-step RSB solution which appears as a finite 'step' in the functions $Q(x)$ and $R(x)$. The entropy of this solution will be shown to be positive.

In the one-step RSB approximation one takes the matrices $\hat{R}$ and $\hat{Q}$ in the form shown in figure 3:

$$R_{\rho\gamma} = r^{\alpha_1\alpha_2}_{\beta_1\beta_2} = \begin{cases} r_0 & \text{if } \alpha_1 \neq \beta_1 \\ r_1 & \text{if } \alpha_1 = \beta_1 \end{cases} \tag{59a}$$

$$Q_{\rho\gamma} = q^{\alpha_1\alpha_2}_{\beta_1\beta_2} = \begin{cases} q_0 & \text{if } \alpha_1 \neq \beta_1 \\ q_1 & \text{if } \alpha_1 = \beta_1 \end{cases} \tag{59b}$$

where $\alpha_1, \beta_1 = 1, \ldots, n/k$ and $\alpha_2, \beta_2 = 1, \ldots, k$.

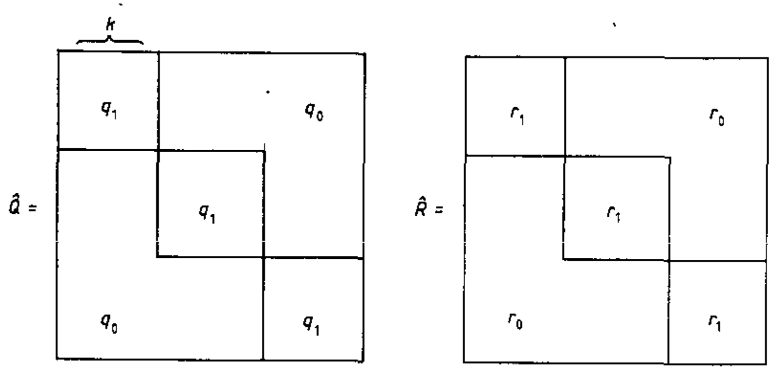

Figure 3.

The free energy depends on r_0, r_1, q_0, q_1 and k, and these parameters should be defined by the corresponding saddle-point equations.

The calculation of the free energy (equation (11)) with matrices $\hat{R}$ and $\hat{Q}$ taken in the form (59) yields (see appendix 2)

$$f=\tfrac{1}{2}a^2+\frac{\alpha}{2}Cr_0+\frac{\alpha}{2}xrq_1+\frac{\alpha}{2x}\ln(1-C)-\frac{\alpha}{2}\frac{q_1}{(1-C)}+\frac{\alpha}{2x}\frac{C}{1-C}$$
$$-\frac{1}{x}\int\frac{\mathrm{d}z}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\ln\left[\int\frac{\mathrm{d}z_1}{\sqrt{2\pi}}\exp\left(-\frac{z_1^2}{2}-\frac{\lambda x}{2}\right.\right.$$
$$\left.\left.\times(a+\sqrt{\alpha r}z_1+\sqrt{\alpha r_0}z)^2\right)[\cosh\beta(a+\sqrt{\alpha r}z_1+\sqrt{\alpha r_0}z)]^{x/\beta}\right] \tag{60}$$

where instead of q_0, k and r_1 we have introduced the parameters

$$x=\beta k$$
$$C=\beta k(q_1-q_0) \tag{61}$$
$$r=r_1-r_0.$$

We study the region $T\ll 1$ and $\alpha\ll T^2$, and therefore in equation (60) the expansion over small α and $\gamma\equiv\exp(-2/T)$ can be made. To first order in γ it yields

$$f(a,q_1,C,r',x)$$
$$\simeq\tfrac{1}{2}(1+\lambda)a^2-a+\frac{\alpha}{2}Cr_0+\frac{\alpha}{2}q_1r'+\frac{\alpha}{2x}\ln(1-C)-\frac{\alpha}{2}\frac{q_1}{(1-C)}+\frac{\alpha}{2x}\frac{C}{1-C}$$
$$+\frac{\alpha}{2}\frac{\lambda r_0}{1+\lambda\alpha r'}+\frac{1}{2x}\ln(1+\lambda\alpha r')-\frac{\alpha}{2}\frac{(1-\lambda a)^2r'}{1+\lambda\alpha r'}$$
$$-T\gamma\exp\left(\frac{2\alpha\beta^2r'}{x(1+\lambda\alpha r')}+\frac{2\alpha\beta^2r_0}{(1+\lambda\alpha r')^2}-\frac{2\alpha\beta(1-\lambda a)r'}{(1+\lambda\alpha r')}\right) \tag{62}$$

where we have introduced $r'=rx$.

The saddle-point equations are

$$a=\frac{1}{1+\lambda} \tag{63}$$

$$r' = \frac{1}{(1-C)} \tag{64}$$

$$r_0 = \frac{q_1}{(1-C)^2} - \frac{1}{x}\frac{C}{(1-C)^2} \tag{65}$$

$$C = -\frac{\lambda}{\Delta} + \frac{4\gamma\beta}{\Delta^2} E \tag{66}$$

$$C = -(1-C)\ln(1-C) - \lambda + \frac{\alpha}{2}\frac{\lambda^2}{(1-C)} + \frac{4\gamma\beta}{\Delta} E \tag{67}$$

$$q_1 = \frac{1}{\Delta}(1-\lambda a)^2 - \frac{\lambda}{\Delta x} + \frac{\alpha\lambda^2 r_0}{\Delta^2} - \frac{\alpha\lambda(1-\lambda a)}{\Delta^2(1-C)} + \left(\frac{4\gamma\beta}{\Delta x} - \frac{4\gamma(1-\lambda a)}{\Delta} - \frac{4\alpha\gamma\beta\lambda}{\Delta^2 x(1-C)} + \frac{4\alpha\gamma\lambda(1-\lambda a)}{\Delta^2(1-C)}\right) E \tag{68}$$

where

$$\Delta = 1 + \lambda\alpha r' \tag{69}$$

$$E = \exp\left(\frac{2\alpha\beta^2 r'}{x\Delta} + \frac{2\alpha\beta^2 r_0}{\Delta^2} - \frac{2\alpha\beta(1-\lambda a) r'}{\Delta}\right). \tag{70}$$

In deriving equations (67) (which is $\partial f/\partial x = 0$) and (68) (which is $\partial f/\partial r' = 0$)), equations (64) have already been used, as well as the approximation

$$\ln(1+\lambda\alpha r') \simeq \lambda\alpha r' - \tfrac{1}{2}(\alpha\lambda r')^2. \tag{71}$$

Solving equations (66) and (67) to leading order in α and λ one obtains

$$C \simeq -\frac{\alpha}{2}\lambda^2. \tag{72}$$

Equation (64) gives

$$r' = 1 - \frac{\alpha}{2}\lambda^2. \tag{73}$$

Then, to leading order in α and λ one obtains from equation (66)

$$4\gamma\beta \exp\left(2\frac{\alpha\beta^2}{x}\right) \simeq \lambda \tag{74}$$

or

$$x \simeq \frac{2\alpha\beta^2}{\ln(\lambda/4\beta\gamma)}. \tag{75}$$

The solution for x exists only if

$$4\gamma\beta < \lambda \tag{76}$$

or

$$T < T^* \simeq \frac{2}{\ln(4/\lambda)}. \tag{77}$$

The solution for the other parameters (to leading order) are

$$\begin{aligned} q_1 &\simeq (1-\lambda)^2 - \lambda T(1-\lambda) \\ q_0 &\simeq q_1 + \tfrac{1}{4}\lambda^2 T^2 \ln\left(\frac{\lambda}{4\beta\gamma}\right) \\ r_0 &\simeq q_1 + \tfrac{1}{4}\lambda^2 T^2 \ln\left(\frac{\lambda}{4\beta\gamma}\right) \\ r_1 &\simeq r_0 + \frac{T^2}{2\alpha} \ln\left(\frac{\lambda}{4\beta\gamma}\right). \end{aligned} \tag{78}$$

The obtained solutions make sense only for $k \leqslant 1$ or $x \leqslant \beta$. To define the line $T_{\mathrm{RSB}}(\alpha)$ below which the solution appears one should take $x = \beta$ in equation (66), to obtain

$$\exp[2\alpha\beta + 2\alpha\beta^2 r_0 - 2\alpha\beta(1-\lambda)] \simeq \frac{\Delta^2}{4\beta\gamma}\left(\frac{\lambda}{\Delta} - \frac{\alpha}{2}\lambda^2\right). \tag{79}$$

To leading order in α, T and λ one obtains

$$T_{\mathrm{RSB}}(\alpha) \simeq T^* - \alpha \tag{80}$$

where T^* is given by equation (77) and $\alpha \ll T^2$. To obtain the entropy the leading terms of the free-energy (62) can be represented as follows:

$$-f \simeq -f_0 + T \exp\left(-\frac{2}{T} + \frac{2\alpha}{xT^2}\right). \tag{81}$$

Then for the entropy one obtains

$$S = -\frac{\partial f}{\partial T} = \frac{\lambda T}{4}\left(1 + \frac{4}{T^*} - \frac{2}{T} + \frac{4\alpha}{T^*}\right). \tag{82}$$

Therefore the entropy is positive below T^* but is becoming negative below $T_0(\alpha)$, where

$$T_0(\alpha) \simeq \tfrac{1}{2}T^* - 2\alpha. \tag{83}$$

Actually the fact that the entropy is becoming negative at low temperatures may represent not the problem of the one-step RSB solution itself, but the problem of the approximation used in deriving the free energy (62) from the exact expression (60). The structure of the obtained solutions (75) is such that the expansion in γ used when deriving the free energy (62) appears to be invalid at $T < T^*/2$. This is also seen from the results for the spin glass order parameter $Q^{\rho\gamma} \equiv \langle \sigma^\rho \sigma^\gamma \rangle$. The physical meaning of the order parameter $\hat{Q}$ (equation (12)) used in all the above calculations is not quite clear. For the physical spin-spin order parameter $\langle \sigma^\rho \sigma^\gamma \rangle$, which has a similar one-step RSB structure, one obtains

$$\langle \sigma_{\alpha_1\alpha_2} \sigma_{\beta_1\beta_2} \rangle|_{\alpha_2 \neq \beta_2} \equiv Q_1 = \overline{\langle \tanh^2 \beta(a + \sqrt{\alpha r_0} z + \sqrt{\alpha r} z_1) \rangle_{z_1}} \tag{84}$$

$$\langle \sigma_{\alpha_1\alpha_2} \sigma_{\beta_1\beta_2} \rangle|_{\alpha_1 \neq \beta_1} \equiv Q_0 = \overline{\langle \tanh \beta(a + \sqrt{\alpha r_0} z + \sqrt{\alpha r} z_1) \rangle_{z_1}^2} \tag{85}$$

where $\overline{(\ldots)}$ denotes the Gaussian average over z and $\langle \ldots \rangle_{z_1}$ indicates the average over

z_1 with the weight

$$\exp\left(-\frac{z_1^2}{2}-\frac{\lambda x}{2}(a+\sqrt{\alpha r_0}z+\sqrt{\alpha r}z_1)^2\right)[\cosh\beta(a+\sqrt{\alpha r_0}z+\sqrt{\alpha r}z_1)]^{x/\beta}. \tag{86}$$

Using the solutions (75) and (78) for the value of the step in the physical order parameter one gets

$$Q_1-Q_0\simeq\gamma^2\left(\frac{\lambda}{4\beta\gamma}\right)^4. \tag{87}$$

This step has a small but finite value everywhere in the region $T<T_{\mathrm{RSB}}(\alpha)$ where the RSB solution exists.

However, the result (87) makes sense only until $\sqrt{\gamma}\gg\lambda$, i.e. for temperatures $T\geqslant T^*/2$. Otherwise, the expansion used when deriving equation (87) as well as all the other results, is invalid.

The results (75) and (87) show that after crossing the transition line $T_{\mathrm{RSB}}(\alpha)$ (equation (80)), the step in the function $Q(x)$ appears at the point $k=1$ where the value of the step is already finite (of the order of $\exp(-1/T)$) and then as $\alpha\to 0$ the position of the step $k\to 0$, thus recovering the RS solution at $\alpha=0$.

The results obtained are schematically shown in figures 4 and 5.

Note that, unlike the results obtained by Krauth and Mezard (1989), the region where the one-step RSB solution exists does not coincide with that defined by the zero-entropy line $T_{S=0}(\alpha)$ of the RS solution.

When lowering the temperature the RSB solution appears *before* the entropy of the RS solution changes sign.

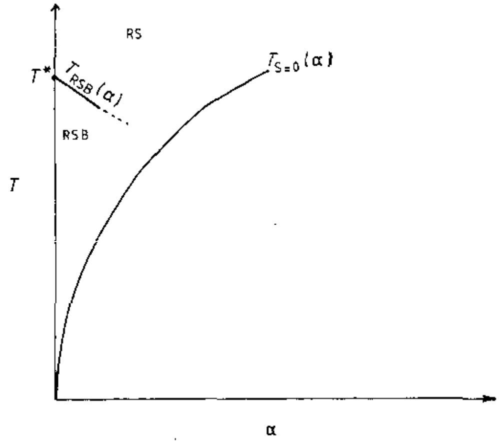

Figure 4.

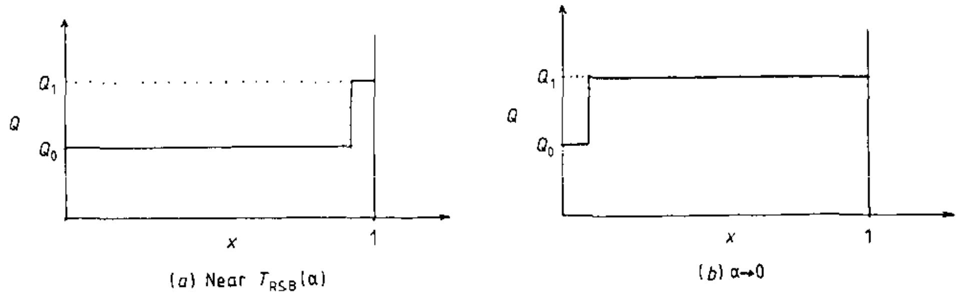

Figure 5.

6. Conclusions

Quite a few questions remain to be answered.

The first is in which way the obtained 'exotic' as well 'normal' RSB solutions in neural networks could be observed, at least in computer simulations. Also, how strongly does the RSB influence, for example, the maximal capacity and the quality of retrieval? The model studied here and in the previous paper (Dotsenko *et al* 1991) could be a good tool for the investigation of these problems. It provides a smooth transition from the Hopfield model ($\lambda = 0$) where the RSB is weak into a situation where the RSB could be expected to be a dominant phenomenon (at λ finite). Note that since the 'exotic' RSB was observed at $\alpha \leqslant 1/\ln(1/\lambda)$ and $T \leqslant 1/\ln(1/\lambda)$ for λ finite, it should be present everywhere in the low temperature (memory) phase of the system.

On the other hand, preliminary results of computer simulations made for this model by Yarunin (1991) show that at least for the critical capacity $\alpha_c(\lambda)$ the experimental curve and that obtained from the RS solution (Dotsenko *et al* 1991) do not differ very much even at λ finite.

There are also more general problems. The present study as well the results obtained by Krauth and Mezard (1989) indicate that in the situation when the entropy of the RS solution changes sign before this solution becomes unstable, the one-step RSB solution may appear to be exact (which is the case for example, in the p-spin interaction SK model for $p \to \infty$ (Gross and Mezard 1984) and in the Potts glass (Gross *et al* 1985)). To what extent the considered phenomenon is general, and what kind of general property allows it to exist in different spin glass-like statistical systems are problems to be solved.

Acknowledgments

One of the authors (VD) would like to thank M Virasoro, G Parisi and M Mezard for fruitful discussions. He is also grateful for the hospitality shown by the Mathematical Department of Rome University I.

Appendix 1

In the exact expression for the free energy (11) we have to put the matrix $\hat{R}$ in the form given by equation (38), then make an expansion over $\hat{r}$ up to fourth order, and then take the limit $n \to 0$.

The limit $n \to 0$ in the second and third terms of the free energy (11) can be taken in a standard way (e.g. see Amit *et al* 1987, Parisi 1979):

$$\frac{\alpha}{2n} \operatorname{tr} \log(\hat{1} - \beta\hat{Q}) \to \lim_{n\to 0} \left(-\frac{\alpha}{2} \int \frac{dx}{x^2} \ln(1-\chi(x)) + \frac{\alpha}{2n} \ln(1-\chi(x) - \beta n Q(n)) \right) \tag{A1.1}$$

$$\frac{\alpha\beta^2}{2n} \sum_{\rho,\gamma} Q_{\rho,\gamma} R_{\rho,\gamma} \to \frac{\alpha\beta^2}{2} Q_0 R_0 - \frac{\alpha\beta^2}{2} \int_0^1 dx\, Q(x) \left(R + \frac{1}{\alpha\beta^2} r(x) \right) \tag{A1.2}$$

where

$$\chi(x) = \beta \left(Q_0 - xQ(x) - \int_x^1 dy\, Q(y) \right). \tag{A1.3}$$

Taking the matrix $\hat{R}$ in the form given by equation (38) one obtains, after some algebra,

$$f=\frac{1}{2}\left(1+\frac{\lambda}{\Delta}\right)a^2+\lim_{n\to 0}\left(-\frac{\alpha}{2}\int_n^1\frac{dx}{x^2}\ln(1-\chi(x)+\frac{\alpha}{2n}\ln(1-\chi(x)-\beta nQ(n))\right)$$
$$+\frac{\alpha\beta}{2}RQ_0-\frac{\Delta-1}{2\lambda}Q_0+\frac{\lambda\alpha R}{2\Delta}-\frac{\Delta-1}{2\lambda\Delta}+\frac{\ln\Delta}{2\beta}$$
$$-\frac{1}{\beta}\overline{\ln\left(\cosh\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)\right)}-\frac{\alpha\beta}{2}\int_0^1 dx\,Q(x)\left(R+\frac{1}{\alpha\beta^2}r(x)\right)$$
$$-\lim_{n\to 0}\frac{1}{\beta n}\ln\left\langle\exp\left(\tfrac{1}{2}\sum_{\rho\neq\gamma}r_{\rho\gamma}(\sigma^\rho+i\lambda\Phi^\rho)(\sigma^\gamma+i\lambda\Phi^\gamma)\right)\right\rangle \tag{A1.4}$$

where the average $\langle(\ldots)\rangle$ in the last term is taken over z, σ and Φ with the weight

$$\exp\left(-\frac{z^2}{2}-\tfrac{1}{2}\lambda\beta\Delta(\Phi-\Phi_0(\sigma))^2+\frac{\beta}{\Delta}(a+\sqrt{\alpha R}z)\sum_\rho\sigma_\rho\right) \tag{A1.5}$$

and

$$\Phi_0(\sigma,z)=\frac{i}{\Delta}\left(a+\sqrt{\alpha R}z+\frac{\Delta-1}{\lambda}\sigma\right). \tag{A1.6}$$

The expansion of the last term in equation (A1.4) up to fourth order gives

$$\ln\left\langle\exp\left(\tfrac{1}{2}\sum_{\rho\neq\gamma}r_{\rho\gamma}(\sigma^\rho+i\lambda\Phi^\rho)(\sigma^\gamma+i\lambda\Phi^\gamma)\right)\right\rangle$$
$$\simeq\tfrac{1}{2}\overline{\langle S\rangle^2}\sum_{\rho\neq\gamma}r_{\rho\gamma}+\tfrac{1}{4}\overline{\langle\langle S^2\rangle\rangle}\sum_{\rho\neq\gamma}r^2_{\rho\gamma}+\tfrac{1}{2}\overline{\langle S\rangle^2\langle\langle S^2\rangle\rangle}\sum_{\rho\neq\gamma,\rho'\neq\gamma}r_{\rho\gamma}r_{\gamma\rho'}$$
$$+\tfrac{1}{6}\overline{\langle\langle S^2\rangle\rangle^3}\,\mathrm{tr}\,\hat{r}^3+\tfrac{1}{12}\overline{\langle\langle S^2\rangle\rangle^3}\sum_{\rho\neq\gamma}r^3_{\rho\gamma}+\tfrac{1}{48}\overline{\langle\langle S^4\rangle\rangle^2}\sum_{\rho\neq\gamma}r^4_{\rho\gamma}. \tag{A1.7}$$

In the RHS of (A1.7) the variable S has been introduced by the definition

$$S_\rho\equiv\sigma^\rho+i\lambda\Phi^\rho.$$

Here as usual to the third and fourth order only the terms which are responsible for the RSB are kept. The notation $\langle\langle(\ldots)\rangle\rangle$ indicates the irreducible average over σ and Φ, and $\overline{(\ )}$ indicates the Gaussian average over z.

In the limit $n\to 0$ one gets

$$\sum_{\rho\neq\gamma}r^k_{\rho\gamma}\to-\int_0^1 dx\,r^k(x)$$
$$\frac{1}{n}\mathrm{tr}\,\hat{r}^3\to\int_0^1\left(xr(x)^3+3r(x)\int_0^x dy\,r^2(y)\right) \tag{A1.8}$$
$$\frac{1}{n}\sum_{\rho\neq\gamma,\rho'\neq\gamma}r_{\rho\gamma}r_{\gamma\rho'}\to-\int_0^1 r^2(x)+\int_0^1 dx\left(xr^2(x)+\int_0^x dy\,r^2(y)+2r(x)\int_0^x dy\,r(y)\right).$$

From equations (A1.8), (A1.7) and (A1.4) one gets the free energy (equation (39)).

Appendix 2

Using the representation of the matrices $\hat{R}$ and $\hat{Q}$ (equation (59)) one easily gets

$$\lim_{n\to 0}\left(\frac{1}{n}\sum_{\rho\gamma} R_{\rho\gamma}Q_{\rho\gamma}\right)=k(q_1r_1-q_0r_0) \tag{A2.1}$$

$$\lim_{n\to 0}\left(\frac{1}{n}\operatorname{tr}\ln(\hat{1}-\beta\hat{Q})\right)=\frac{1}{k}\ln[1-\beta k(q_1-q_0)]-\frac{\beta q_0}{1-\beta k(q_1-q_0)}. \tag{A2.2}$$

For the term $\Sigma_{\rho\gamma}R_{\rho\gamma}S_\rho S_\gamma$ one gets

$$\begin{aligned}\sum_{\rho\gamma} R_{\rho\gamma}S_\rho S_\gamma &= r_1\sum_{\alpha_1=1}^{n/k}\sum_{\alpha_2\beta_2=1}^{k} S_{\alpha_1\alpha_2}S_{\alpha_1\beta_2}+r_0\sum_{\alpha_1\neq\beta_1}^{n/k}\sum_{\alpha_2\beta_2=1}^{k} S_{\alpha_1\alpha_2}S_{\beta_1\beta_2}\\ &=(r_1-r_0)\sum_{\alpha_1=1}^{n/k}\left(\sum_{\alpha_2=1}^{k} S_{\alpha_1\alpha_2}\right)^2+r_0\left(\sum_{\alpha_1\alpha_2=1}^{k} S_{\alpha_1\alpha_2}\right)^2.\end{aligned} \tag{A2.3}$$

Therefore, the last term of the free energy (11) can be represented as follows:

$$\begin{aligned}-\frac{1}{\beta n}\ln\Bigg\{\int\frac{\mathrm{d}z}{\sqrt{2\pi}}\left(\prod_{\alpha_1=1}^{n/k}\int\frac{\mathrm{d}z_{\alpha_1}}{\sqrt{2\pi}}\right)\sum_{(\sigma\rho)}\int \mathrm{D}\Phi\exp\Bigg[-\frac{z^2}{2}-\tfrac{1}{2}\sum_{\alpha_1=1}^{n/k} z_{\alpha_1}^2\\ +\sum_{\alpha_1=1}^{n/k}\sum_{\alpha_2=1}^{k}\left(-\frac{\lambda\beta}{2}\Phi_{\alpha_1\alpha_2}^2+\beta\sqrt{\alpha(r_1-r_0)}S_{\alpha_1\alpha_2}z_{\alpha_1}\right.\\ \left.+\beta\sqrt{\alpha r_0}S_{\alpha_1\alpha_2}z+\beta a S_{\alpha_1\alpha_2}\right)\Bigg]\Bigg\}.\end{aligned} \tag{A2.4}$$

After the definition of the variable $\Phi\to(1/\mathrm{i}\lambda)\sigma+\tilde{\Phi}$ and integrating over $\tilde{\Phi}$ one gets

$$\begin{aligned}-\frac{1}{2\lambda}-\frac{1}{\beta n}\ln\int\frac{\mathrm{d}z}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\Bigg\{\int\frac{\mathrm{d}z_1}{\sqrt{2\pi}}\exp\left(-\frac{z_1^2}{2}\right)\\ \times\left[\sum_{\sigma=\pm1}\exp\left(-\frac{\beta}{2\lambda}\{\sigma+\lambda[a+\sqrt{\alpha(r_1-r_0)}z_1+\sqrt{\alpha r_0}z]^2\}\right)\right]^k\Bigg\}^{n/k}.\end{aligned} \tag{A2.5}$$

In the limit $n\to 0$ one obtains

$$\begin{aligned}-\frac{1}{\beta k}\int\frac{\mathrm{d}z}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\ln\Bigg[\int\frac{\mathrm{d}z_1}{\sqrt{2\pi}}\exp\left(-\frac{z_1^2}{2}-\frac{\beta\lambda k}{2}[a+\sqrt{\alpha(r_1-r_0)}z_1+\sqrt{\alpha r_0}z]^2\right)\\ \times\cosh^k\beta(a+\sqrt{\alpha(r_1-r_0)}z_1+\sqrt{\alpha r_0}z)\Bigg].\end{aligned} \tag{A2.6}$$

Equations (A2.1), (A2.2) and (A2.6) give the result (60).

References

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

de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983

Dotsenko V S, Yarunin N D and Dorotheyev E A 1991 *J. Phys. A: Math. Gen.* **24** 2419

Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271

Gross D J and Mézard M 1984 *Nucl. Phys.* B **240** 431

Gross D J, Kanter I and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 304

Gutfreund M and Stein Y 1990 *J. Phys. A: Math. Gen.* **23** 2613

Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
Krauth W and Mézard M 1989 *J. Phys.* **50** 3057
Mézard M, Parisi G and Virasoro M 1987 *Spin-Glass Theory and Beyond* (Singapore: World Scientific)
Parisi G 1979 *Phys. Rev. Lett.* **43** 1754
Parisi G 1980 *J. Phys. A: Math. Gen.* **13** 1887
Personaz L, Guyon I and Dreyfus G 1985 *J. Phys. Lett.* **46** L359
Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **32** 1972
Yarunin N 1991 unpublished