Stability of the one-step replica-symmetry-broken phase in neural networks

1992 J. Phys. A: Math. Gen. 25 5527

(http://iopscience.iop.org/0305-4470/25/21/012)

J. Phys. A: Math. Gen. 25 (1992) 5527–5537. Printed in the UK

Stability of the one-step replica-symmetry-broken phase in neural networks

E A Dorotheyev

Landau Institute for Theoretical Physics, 142432 Chernogolovka, Moscow region, Russia

Received 24 September 1991, in final form 23 April 1992

Abstract. The stability of the phase with one-step 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 temperature there is a region of parameters where the one-step replica-symmetry-broken solution exists. Fluctuations around this solution are analysed by a replica group representations approach and it is found that the solution is stable for all ranges of the parameters where it exists.

1. Introduction

In the studies of the phenomenon of replica symmetry breaking (RSB), the problem of the one-step replica symmetry breaking (1RSB) in the spin-glass-like statistical systems is of a special interest. Usually, 1RSB is considered as a first approximation in the Parisi (1980) scheme for the description of the fully replica symmetry broken low temperature phase. A classical example is the Sherrington and Kirkpatrick (1975) model of a spin-glass for which the phenomenon of RSB is already well studied (Mezard *et al* 1987).

On the other hand, there are several systems for which it is known that 1RSB gives the exact solution. Among them are: the p-spin interactions ($p \to \infty$) version of the SK problem (Gross and Mezard 1984), and the Potts spin glasses (Gross *et al* 1985, Cwilich and Kirkpatrick 1989). Recently, a similar phenomenon was discovered by Dotsenko and Tirozzi (1991) in a model of a neural network.

It is commonly believed that the RSB phenomenon is not very important for neural networks, in contrast to spin-glasses where RSB is a crucial characteristic of the low-temperature phase. A classical example is the model proposed by Hopfield (1982). It was shown (see, for example, Amit *et al* 1987) that although at sufficiently low temperature the replica-symmetric (RS) solution is unstable against RSB, the instability is very weak and the true RSB solution is not much different from the RS one.

For that model the phenomenon of RSB itself was proved 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) (1978).

However the RSB phenomenon in the model of neural networks considered by Dotsenko and Tirozzi (1991) appears to be of qualitatively different kind. They showed that there is a certain range of parameters where the RS solution is still stable, but its entropy is negative (so that the zero-entropy line goes above the AT-line). It indicates that in addition to the stable RS solution there exists the other RSB solution, which 'absorbs' a part of the entropy. It was shown that the structure of this RSB state is essentially different from the 'traditional' one in the SK and the Hopfield models, since it does not appear as a result of instability of the RS solution. The corresponding RSB state was obtained to be 1RSB one, although its stability remained an open question.

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

Therefore it would be reasonable to expect that such 'exotic' 1RSB solutions might be a rather general phenomenon for a certain class of SG-like systems, and it makes sense to investigate the stability of the 1RSB phase in a model of neural networks.

The plan of the paper is the following. In section 2 the model is introduced and its mean-field free energy is obtained. In section 3 the 1RSB solution is presented. In section 4 fluctuations around this solution are considered and the stability criteria is derived. In section 5 the stability of the 1RSB solution is proved.

2. The model

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

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

The interaction matrix is taken to be of 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

$$C_{\mu\nu}=\frac{1}{N}\sum_i \xi_i^\mu\xi_j^\nu \tag{3}$$

and ξ_i^μ $(\mu=1,\ldots,p)$ are quenched uncorrelated patterns, λ is the parameter of the model. If $\lambda=0$ the model (1), (2) turns into the Hopfield one, and as $\lambda\to\infty$ the structure of interactions matrix (2) approaches that of the pseudo-inverse model studied by Personaz *et al* (1985) and by Kanter and Sompolinsky (1987). The motivation to this particular choice of the J_{ij} has been given by Dotsenko *et al* (1991). It could be obtained from the traditional Hebb learning rule via local thermally noised iterative procedure, and the corresponding RS solution provides a substantial increase of the capacity and of the 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 $\langle\langle\ldots\rangle\rangle$ means the 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^ρ_μ, Φ^ρ_i one gets

$$\langle\langle Z^n\rangle\rangle = \int \mathrm{D}\underline{a} \int \mathrm{D}\underline{\Phi} \sum_{\{\sigma_i^\rho\}} \exp\biggl(-\frac{\beta}{2N}\sum_{\rho\mu}(a^\rho_\mu)^2 - \tfrac{1}{2}\sum_i (\Phi^\rho_i)^2 + \beta \sum_{i\mu\rho} a^\rho_\mu \xi^\mu_i \biggl(\sigma^\rho_i + \mathrm{i}\sqrt{\frac{\lambda}{\beta}}\,\Phi^\rho_i\biggr)\biggr) \tag{6}$$

where

$$\mathrm{D}\underline{a} = \prod_{\mu\rho} \mathrm{d}a^\rho_\mu \qquad \mathrm{D}\underline{\Phi} = \prod_{i\rho} \frac{\mathrm{d}\Phi^\rho_i}{\sqrt{2\pi}}.$$

The irrelevant factor containing $\det(\hat{1} + \lambda\hat{C})$ is omitted. The fields a_μ are connected with the usual overlaps

$$m_\mu = \frac{1}{n}\sum_i \xi^\mu_i \langle\sigma_i\rangle \tag{7}$$

as follows

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

By standard calculations similar to those for the Hopfield model (Amit *et al* 1987) after averaging over the ξ^μ_i one gets

$$\langle\langle Z^n\rangle\rangle = \int \mathrm{D}\underline{a} \int \mathrm{D}\underline{Q} \int \mathrm{D}\underline{R}\, \exp(-\beta N n f(\underline{a}, \underline{Q}, \underline{R})) \tag{9}$$

where

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

Here f is the mean field free energy of the model:

$$-\beta N n f(\underline{a}, \underline{Q}, \underline{R}) = -\frac{\beta N}{2}\sum_\rho (a^\rho)^2 - \frac{\alpha N}{2}\mathrm{Tr}\log(\hat{1} - \beta\hat{Q}) + \frac{\mathrm{i}N}{2}\sum_{\rho\gamma} R_{\rho\gamma} Q_{\rho\gamma} + N \log \mathrm{Tr}_{\sigma\phi} \exp(-\beta F) \tag{10}$$

$$-\beta F = \beta\sum_\rho a^\rho \biggl(\sigma^\rho + \mathrm{i}\sqrt{\frac{\lambda}{\beta}}\,\Phi^\rho\biggr) - \frac{\mathrm{i}}{2}\sum_{\rho\gamma} R_{\rho\gamma}\biggl(\sigma^\rho + \mathrm{i}\sqrt{\frac{\lambda}{\beta}}\,\Phi^\rho\biggr)\biggl(\sigma^\gamma + \mathrm{i}\sqrt{\frac{\lambda}{\beta}}\,\Phi^\gamma\biggr) \tag{11}$$

$$\mathrm{Tr}_{\rho\phi}(\ldots) = \sum_{\{\sigma^\rho\}} \prod_\rho \int_{-\infty}^{\infty} \frac{\mathrm{d}\Phi^\rho}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(\Phi^\rho)^2\}(\ldots). \tag{12}$$

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

$$Q_{\rho\gamma} = \frac{1}{N}\sum_i \biggl(\sigma^\rho_j + \mathrm{i}\sqrt{\frac{\lambda}{\beta}}\,\Phi^\rho_j\biggr)\biggl(\sigma^\gamma_i + \mathrm{i}\sqrt{\frac{\lambda}{\beta}}\,\Phi^\gamma_j\biggr) \tag{13}$$

and the matrix $R_{\rho\gamma}$ has been defined as a conjugate field to equation (13).

In the above calculations the pattern number 1 (i.e. $\{\xi^1_i\}$) was assumed to be condensed and therefore the parameter a^ρ in (10) has been defined as

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

3. One-step replica-symmetry-broken solution

Following Dotsenko and Tirozzi (1991) let us try a RSB ansatz for the fields $R_{\alpha\beta}$, $Q_{\alpha\beta}$ as follows. Introduce the two-position parametrization of the replica set: group the n replicas in clusters of k, where

$$\alpha=[\alpha_1,\alpha_2] \qquad \alpha_1=1,\ldots,n/k \qquad \alpha_2=1,\ldots,k \tag{14}$$

where k is a parameter to be fixed by the saddle point equations. Here α_1 enumerates blocks of replicas and α_2 enumerates replicas in the blocks.

Using this parametrization let us define

$$Q_{\alpha\beta}=\begin{cases} q_1 & \text{if } \alpha=\beta_1 \\ q_0 & \text{otherwise} \end{cases} \qquad R_{\alpha\beta}=\begin{cases} r_1 & \text{if } \alpha_1=\beta_1 \\ r_0 & \text{otherwise.} \end{cases} \tag{15}$$

Substituting this ansatz in (10) one obtains

$$\begin{aligned} f=\tfrac{1}{2}a^2+\frac{\alpha}{2}cr_0+\frac{\alpha\beta}{2}krq_1+\frac{\alpha}{2\beta k}\log(1-c)-\frac{\alpha q_1}{2(1-c)}+\frac{\alpha}{2\beta k}\frac{c}{1-c} \\ -\frac{1}{\beta k}\int \mathrm{D}z\log\left[\int \mathrm{D}z_1\exp\left(-\frac{\lambda\beta k}{2}u^2\right)(\cosh(\beta u))^k\right] \end{aligned} \tag{16}$$

where

$$\begin{aligned} u&=a+\sqrt{\alpha r}z_1+\sqrt{\alpha r_0}z \\ \mathrm{D}z&=\frac{\mathrm{d}z}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right) \\ \mathrm{D}z_1&=\frac{\mathrm{d}z_1}{\sqrt{2\pi}}\exp\left(-\frac{z_1^2}{2}\right) \end{aligned} \tag{17}$$

and instead of q_0, r_1 we have introduced

$$c=\beta k(q_1-q_0) \qquad r=r_1-r_0.$$

The solutions of the saddle point equations obtained from the free energy (16) has been found by Dotsenko and Tirozzi (1991). It was shown that below the transition line $T_{\mathrm{RSB}}(\alpha)=T^*-\alpha$, where $T^*\simeq 2/\ln(4/\lambda)$ and $\alpha\ll T^2$, the 1RSB solution can be represented as follows:

$$\begin{aligned} &a\approx 1 \qquad q_1\approx 1 \qquad q_0\approx 1 \qquad r_0\approx 1 \\ &r=\frac{T(T^*-T)}{\alpha T^*} \\ &k=\left\{1+\frac{(T-\alpha)(T_{\mathrm{RSB}}(\alpha)-T)}{\alpha T^* T}\right\}^{-1}. \end{aligned} \tag{18}$$

4. The stability criteria

To investigate the stability of the solution let us introduce the variables

$$a_\alpha=a+\varepsilon_\alpha \qquad Q_{\alpha\beta}=Q^{(0)}_{\alpha\beta}+\kappa_{\alpha\beta} \qquad R_{\alpha\beta}=R^{(0)}_{\alpha\beta}+\rho_{\alpha\beta}$$

where a, $Q^{(0)}_{\alpha\beta}$, $R^{(0)}_{\alpha\beta}$ are the equilibrium values given by equations (15) and (18). Substituting a, $Q_{\alpha\beta}$, $R_{\alpha\beta}$ in (10) and expanding to second order in the fluctuations, we obtain the variation $-\frac{1}{2}\Delta$ of the free energy with

$$\Delta = \sum_{\alpha\beta} H^{aa}_{\alpha\beta}\varepsilon_\alpha\varepsilon_\beta + 2\sum_{(\alpha\beta)\delta} H^{ra}_{(\alpha\beta)\delta}\rho_{\alpha\beta}\varepsilon_\delta + \sum_{(\alpha\beta)(\gamma\delta)} H^{rr}_{(\alpha\beta)(\gamma\delta)}\rho_{\alpha\beta}\rho_{\gamma\delta} - 2\mathrm{i}\sum_{(\alpha\beta)}\rho_{\alpha\beta}\kappa_{\alpha\beta} + \sum_{(\alpha\beta)(\gamma\delta)} H^{qq}_{(\alpha\beta)(\gamma\delta)}\kappa_{\alpha\beta}\kappa_{\gamma\delta} \tag{19}$$

(the notation $(\alpha\beta)$ means that each pair is counted only once). The submatrices $\hat{H}^{aa}$, $\hat{H}^{ra}$, $\hat{H}^{rr}$, $\hat{H}^{qq}$ of the Hessian are given by

$$H^{aa}_{\alpha\beta} = \beta\delta_{\alpha\beta} - \beta^2[\langle s^\alpha s^\beta\rangle - \langle s^\alpha\rangle\langle s^\beta\rangle] \tag{20}$$

$$H^{ra}_{(\alpha\beta)\gamma} = \mathrm{i}\beta[\langle s^\alpha s^\beta s^\gamma\rangle - \langle s^\alpha s^\beta\rangle\langle s^\gamma\rangle] \tag{21}$$

$$H^{rr}_{(\alpha\beta)(\gamma\delta)} = \langle s^\alpha s^\beta s^\gamma s^\delta\rangle - \langle s^\alpha s^\beta\rangle\langle s^\gamma s^\delta\rangle \tag{22}$$

$$H^{qq}_{(\alpha\beta)(\gamma\delta)} = -2\beta(G_{\alpha\delta}G_{\beta\gamma} + G_{\alpha\gamma}G_{\beta\delta}) \tag{23}$$

where

$$s^\alpha = \sigma + \mathrm{i}\sqrt{\frac{\lambda}{\beta}}\,\Phi^\alpha \tag{24}$$

$$\hat{G} = (\hat{1} - \beta\hat{Q})^{-1} \tag{25}$$

$$\langle\ldots\rangle = \frac{\mathrm{Tr}_{\sigma\phi}(\ldots)\exp(-\beta F)}{\mathrm{Tr}_{\sigma\phi}\exp(-\beta F)}. \tag{26}$$

The submatrices (20)-(23) should be calculated with $R_{\alpha\beta}$, $Q_{\alpha\beta}$ taken in the form (15) and F, $\mathrm{Tr}_{\sigma\phi}$ in (26) is given by (11), (12) respectively.

For the steepest descent method to be correct, it is necessary that the integral

$$\int \mathrm{D}\underline{\varepsilon}\,\mathrm{D}\underline{\kappa}\,\mathrm{D}\underline{\rho}\,\exp(-\tfrac{1}{2}\Delta) \tag{27}$$

should converge in the vicinity of the saddle point.

There are two problems, we have to focus on.

The first is to evaluate the matrix elements of the submatrices (20)-(23). To do this, one should define different types of matrix elements. For example, in the RS case, the submatrices $\hat{H}^{qq}$ have five types of matrix elements namely $H^{qq}_{(1,1)(1,1)}$, $H^{qq}_{(1,1)(2,2)}$, $H^{qq}_{(1,2)(1,2)}$, $H^{qq}_{(1,2)(1,3)}$, $H^{qq}_{(1,2)(3,4)}$. The 1RSB case is more complicated.

The second is to reduce the quadratic form (19) to the quadratic form of the simplest type by admissible coordinate transformations.

A reduction of the quadratic form (19) to the quadratic form of simplest type by admissible coordinate transformation in general cannot be reduced to a diagonalization of matrix

$$\begin{pmatrix} \hat{H}^{aa} & \hat{H}^{ar} & 0 \\ \hat{H}^{ar} & \hat{H}^{rr} & -\mathrm{i}\hat{1} \\ 0 & -\mathrm{i}\hat{1} & \hat{H}^{qq} \end{pmatrix}.$$

This is a *complex* matrix, but linear complex orthogonal transformations are not admissible for the integral (27). Therefore, one should combine real orthogonal transformations of the coordinates ε_α, $\rho_{\alpha\beta}$, $\kappa_{\alpha\beta}$ with a subsequent proper shift of the integration contour in complex plane.

With $Q_{\alpha\beta}$, $R_{\alpha\beta}$ taken in the form (15), the submatrices (20)–(23) have a complicated structure, therefore these two problems are non-trivial. A simplification occurs if one takes into account the fact that the submatrices are left invariant under the action of some permutation group. This group may be called the hierarchical tree (HT) group since it conserved the structure of the hierarchical tree connected with ansatz (15). The HT group is described in appendix 1.

Different types of matrix elements can be obtained by consideration of HT group orbits in the sets $\{(\alpha, \beta), (\gamma, \delta)\}$, $\{(\alpha, (\beta, \gamma))\}$. These matrix elements are presented in appendix 2.

The simplification of the quadratic form (19) can be achieved by choosing bases of the subspaces $\{\varepsilon_\alpha\}$, $\{\rho_{\alpha\beta}\}$, $\{\kappa_{\alpha\beta}\}$ to be the bases of irreducible representations of the HT group. The other words, these bases are distinct families of fluctuation modes. Decomposition of the subspaces in irreducible components is described in appendix 3.

It is well known (e.g. see Almeida and Thouless 1979, Amit *et al* 1987), that only the replicon-like families of modes are dangerous in the sense of the stability, and they are denoted as $R^{(a)}$, $R_1^{(e)}$, $R_2^{(e)}$, $R_3^{(e)}$ in appendix 3. The vectors of these families have non-zero eigenvalues only for the submatrices

$$H^{qq}_{(\alpha\beta)(\gamma\delta)} \qquad H^{rr}_{(\alpha\beta)(\gamma\delta)} \qquad (\alpha \neq \beta;\ \gamma \neq \delta).$$

Using the information from appendices 1, 2 and table 1 these eigenvalues can be found. They are given by

$$R^{(a)} \to \begin{cases} \lambda^q = -\alpha\beta^2 \\ \lambda^r = \overline{\langle A^2 \rangle_{z_1}} \end{cases} \tag{28}$$

$$R_1^{(e)} \to \begin{cases} \lambda^q = -\dfrac{\alpha\beta^2}{(1 - k\beta(q_1 - q^0))^2} \\ \lambda^r = \overline{(\langle A \rangle_{z_1} + k\langle B^2 \rangle_{z_1} - k\langle B \rangle_{z_1}^2)^2} \end{cases} \tag{29}$$

$$R_2^{(e)} \to \begin{cases} \lambda^q = -\alpha\beta^2 \\ \lambda^r = \overline{\langle A \rangle_{z_1}^2} \end{cases} \tag{30}$$

$$R_3^{(e)} \to \begin{cases} \lambda^q = -\dfrac{\alpha\beta^2}{1 - k\beta(q_1 - q^0)} \\ \lambda^r = \overline{(\langle A \rangle_{z_1} + k\langle B^2 \rangle_{z_1} - k\langle B \rangle_{z_1}^2)\langle A \rangle_{z_1}} \end{cases} \tag{31}$$

where

$$A = \cosh^{-2}(\beta u) - \frac{\lambda}{\beta}$$

$$B = \tanh(\beta u) - \lambda u.$$

Table 1. Eigenvalues of matrices $\hat{H}^{qq}$, $\hat{H}^{rr}$ for each replicon-like family.

Family	Eigenvalue
R^a	$K_1 - 2K_2 + K_3$
R_1^e	$L_1 - 2L_2 + L_3 + 2k(L_2 - L_3 - L_4 + L_5) + k^2(L_3 - 2L_5 + L_6)$
R_2^e	$L_1 - 2L_2 + L_3$
R_3^e	$L_1 - 2L_2 + L_3 + k(L_2 - L_3 - L_4 + L_5)$

In equations (28)–(31) $\overline{(\ldots)}$ means Gaussian average over z, $\langle\ldots\rangle_{z_1}$ means the average over z_1 with the weight

$$\exp\left(-\frac{z_1^2}{2}-\frac{\lambda\beta k u^2}{2}\right)(\cosh(\beta u))^k$$

and u is given by (17).

The contribution of modes of the replicon-like families to the integral (27) may be represented as follows:

$$\prod_{R=R^{(a)},R_1^{(e)},R_2^{(e)},R_3^{(e)}}\left(\int \mathrm{d}r\,\mathrm{d}q\,\exp\left[-\frac{N}{2}(\lambda_R^q q^2-2\mathrm{i}qr+\lambda_R^r r^2)\right]\right)^{\dim_R} \tag{32}$$

where λ^q, λ^r are given by equations (28)–(31), and $\dim_R$ is the dimension of the irreducible representation R.

According to equations (28)–(31) λ_R^q are always negative. Nevertheless, if $\lambda_R^r>0$ for all R the integral (32) appears to be absolutely converging, since according to the definition of the matrix $Q_{\alpha\beta}$ (13) the integration over $\mathrm{D}_{\underline{K}}$ in the integral (27) runs over a finite domain. Consequently, one may integrate over r and over q in any order. By integrations over r in (32) we find

$$\prod_{R=R^a,R_1^e,R_2^e,R_3^e}\left(\int \mathrm{d}q\,\exp\left[-\frac{N}{2}\left(\lambda_R^q+\frac{1}{\lambda_R^r}\right)q^2\right]\right)^{\dim_R}.$$

Hence, the integral (32) is dominated by the vicinity of the saddle point if

$$1+\lambda_R^r\lambda_R^q>0 \qquad \lambda_R^r>0 \tag{33}$$

for all families $R=R^{(a)}, R_1^{(e)}, R_2^{(e)}, R_3^{(e)}$. These inequalities with λ_R^r, λ_R^q from equations (28)–(31) are the stability criteria.

5. The 1RSB phase stability

Let us consider the inequality (33) in the case, that the parameters a, q_0, q_1, r_0, r_1, k are given by (18). At first, we consider the vicinity of the transition line $T_{\mathrm{RSB}}(\alpha)=T^*-\alpha$. In this region $T^*-T\ll T^*$, $k\simeq 1$ and the parameter $\beta\sqrt{\alpha r}$ is small:

$$\beta\sqrt{\alpha r}=\left(\frac{T^*-T}{T^*T}\right)^{1/2}\ll 1.$$

This means that for the average $\langle f(z_1)\rangle_{z_1}$, one has

$$\langle f(z_1)\rangle_{z_1}=f(z_1=0)$$

since the variable z_1 is always multiplied by the factor $\beta\sqrt{\alpha r}$ (with the only exception for $\exp(-z_1^2/2)$). So each inequality (33) is transformed to

$$1-\frac{\alpha}{T_*^2}\int\frac{\mathrm{d}z}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\cosh^{-4}\left(\frac{a+\sqrt{\alpha}\,z}{T^*}\right)>0. \tag{34}$$

Integrating over z, one obtains

$$1-\frac{T_{AT}(\alpha)}{T^*}>0$$

where

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

is the Almeida-Thouless temperature, which defines the boundary of the stability domain of the RS phase (Amit *et al* 1987).

In the region $\alpha \ll T^2$

$$T_{AT}(\alpha)\ll T^*$$

therefore the 1RSB phase is stable in the vicinity of the transition line. The correction to (34) induced by expansion of the stability criteria (equation (33)) in powers of $\beta\sqrt{\alpha r}$, do not change this result.

Now let us consider the case $T\ll T^*$, $\alpha \ll T$. Then

$$\beta\sqrt{\alpha r}\simeq\sqrt{1/T}\gg 1$$

$$k\simeq\alpha\ll 1$$

$$\beta u\equiv\beta(a+\sqrt{\alpha r}\, z_1+\sqrt{\alpha r_0}\, z)\simeq\beta+\sqrt{\beta}\, z_1+\beta\sqrt{\alpha}\, z.$$

The averaging over z may be omitted, because $\beta\sqrt{\alpha}\ll 1$. The averaging over z_1 is reduced to

$$\langle f(\beta u)\rangle_{z_1}=\int\frac{\mathrm{d}z_1}{\sqrt{2\pi}}\exp\left(-\frac{z_1^2}{2}\right)f(\beta+\sqrt{\beta}\, z_1).$$

By this averaging rule, the stability criteria (equation (33)) can be reduced in the leading order in α to

$$1-2\pi\alpha\beta\exp(-\beta)>0 \qquad \text{(for } R^{(e)} \text{ families)}$$

$$1-\frac{9}{4}\sqrt{\frac{\pi}{2}}\,\alpha\beta\exp\left(-\frac{\beta}{2}\right)>0 \qquad \text{(for } R^{(a)} \text{ family).}$$

These inequalities are true if $\alpha\ll T^2$, and therefore the 1RSB phase is stable in the region $T\ll T^*$, $\alpha\ll T^2$.

6. Conclusions

Thus, we have established the stability of the 1RSB phase in the region $T\ll 1$, $\alpha\ll T^2$. A line of the type of the AT-line was not found in this region. It is possible, that this line is located in the region $T\ll\alpha$, where we do not have a solution of the type equation (18).

At the same time, the 1RSB phase is stable below the line $T_0(\alpha)=\frac{1}{2}T^*-2\alpha$, where the entropy of the 1RSB phase becomes negative (Dotsenko and Tirozzi 1991). The analogous phenomenon for the RS phase (the entropy is negative while the solution is still stable) has initiated a search for the 1RSB phase. This phase is believed to absorb a part of the entropy. In our case, if the entropy sign change is not an artifact of the approximation used for the free energy evaluation the following question arises: what is the phase which 'absorbs' part of the entropy? Presumably it may be the two-step RSB (2RSB) phase, and therefore an assumption that in fact we deal with a whole cascade of 'first-order' RSB phase transitions does not seem to be unrealistic.

Acknowledgment

I would like to thank V S Dotsenko for numerous discussions on the work presented here.

Appendix 1

The HT group can be represented as follows (e.g. see Mezard *et al* 1987):

$$S_{n/k}\hat{\otimes}(S_k)^{\otimes n/k}$$

where $(S_k)^{\otimes n/k}$ is the direct product of the permutation group of k elements (S_k) with itself n/k times, and $\hat{\otimes}$ means the semidirect product.

Indeed we can permute both the replicas inside each cluster (it leads to the product of S_k by itself n/k times) and the clusters of replicas among themselves (this leads to $S_{n/k}$).

Any element of the HT group may be written as follows

$$g=(g^{(1)};\, g_1^{(2)},\ldots,g_{n/k}^{(2)})$$

where

$$g^{(1)}\subset S_{n/k}$$

$$g_i^{(2)}\subset S_k \qquad i=1,\ldots,n/k.$$

The action of g on the replica set is given by

$$g\alpha=[g^{(1)}\alpha_1,\, g_{\alpha_1}^{(2)}\alpha_2]$$

where

$$\alpha=[\alpha_1,\alpha_2]$$

and $g^{(1)}\alpha_1, g_{\alpha_1}^{(2)}\alpha_2$ are the results of the action of the groups $S_{n/k}, S_k$ on the sets $\{1,\ldots,n/k\}$ and $\{1,\ldots,k\}$ respectively.

If $h=(h^{(1)};\, h_1^{(2)},\ldots,h_{n/k}^{(2)})$ then

$$gh=(g^{(1)}h^{(1)};\, g_{g^{(1)}1}^{(2)}h_1^{(2)},\ldots,g_{g^{(1)}n/k}^{(2)}h_{n/k}^{(2)}).$$

This composition law means that the HT group is a semidirect product $S_{n/k}$ with $(S_k)^{\otimes n/k}$.

Appendix 2

Independent parameters which define the matrices $\hat{H}^{qq}$, $\hat{H}^{rr}$ are listed below. Only the matrix elements of the replicon-like families relevant for stability are represented. (The notation $H_{(\alpha\beta),(\gamma\delta)}=\langle\alpha\beta|\hat{H}|\gamma\delta\rangle$ and the parametrization of the replica set (14) are used, superscripts (qq), (rr) are omitted.)

Intracluster matrix elements:

$$\langle[1,1],[1,2]|\hat{H}|[1,1],[1,2]\rangle=K_1$$

$$\langle[1,1],[1,2]|\hat{H}|[1,1],[1,3]\rangle=K_2$$

$$\langle[1,1],[1,2]|\hat{H}|[1,3],[1,4]\rangle=K_3$$

$$\langle[1,1],[1,2]|\hat{H}|[2,1],[2,2]\rangle=K_4.$$

Intercluster matrix elements:

$$\langle [1,1],[2,1]|\hat{H}|[1,1],[2,1]\rangle = L_1$$

$$\langle [1,1],[2,1]|\hat{H}|[1,1],[2,2]\rangle = L_2$$

$$\langle [1,1],[2,1]|\hat{H}|[1,2],[2,2]\rangle = L_3$$

$$\langle [1,1],[2,1]|\hat{H}|[1,1],[3,1]\rangle = L_4$$

$$\langle [1,1],[2,1]|\hat{H}|[1,2],[3,1]\rangle = L_5$$

$$\langle [1,1],[2,1]|\hat{H}|[3,1],[4,1]\rangle = L_6.$$

The other matrix elements of $\hat{H}^{qq}$, $\hat{H}^{rr}$ may be obtained from these by the HT group action on indexes.

Appendix 3

We have three different types of subspaces in the space $(\{\varepsilon_\alpha\}, \{\rho_{\alpha\beta}\}, \{\kappa_{\alpha\beta}\})$ of all modes which are not mixed by permutations of the replicas.

The first type: n-dimensional subspaces

$$\{\varepsilon_\alpha\} \qquad \{\rho_{\alpha\alpha}\} \qquad \{\kappa_{\alpha\alpha}\}$$

which may be called diagonal subspaces.

The second type: $n(k-1)/2$-dimensional subspaces

$$\{\rho_{\alpha\beta}\} \qquad \{\kappa_{\alpha\beta}\}$$

with $\alpha \neq \beta$ and α, β belonging to the same cluster. They may be called intracluster subspaces. Irreducible representations in this subspace would be marked by a superscript (a).

The third type: $n(n-k)/2$-dimensional subspaces

$$\{\rho_{\alpha\beta}\} \qquad \{\kappa_{\alpha\beta}\}$$

with $\alpha \neq \beta$ and α, β belonging to the different cluster. They may be called intercluster subspaces. Irreducible representations in this subspace would be marked by a superscript (e).

Any subspace of the first type may be decomposed into a sum of three irreducible representations of the HT group. One of them, L, corresponds to a longitude mode, other A_1, A_2 correspond to anomalous-like families of the modes.

The subspaces of the second type may be decomposed into a sum of four irreducible components. The first three of them are equivalent to L, A_1, A_2, while the remaining one appears to be a replicon-like family $R^{(a)}$.

The subspaces of the third type may be decomposed into a sum of six irreducible components. Three of them coincide with L, A_1, A_2, and three are replicon-like families $R_1^{(e)}$, $R_2^{(e)}$, $R_3^{(e)}$.

The detailed review of their structure, dimensions and characters is to be published elsewhere.

References

Amit D I, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30
Cwilich G and Kirkpatrick 1989 *J. Phys. A: Math. Gen.* **22** 4971
de Almedia J R L and Thouless D J 1987 *J. Phys. A: Math. Gen.* **11** 983
Dotsenko V and Tirozzi B 1991 *J. Phys. A: Math. Gen.* **24** 5163
Dotsenko V, Yarunin N and Dorotheyev E 1991 *J. Phys. A: Math. Gen.* A **24** 2419
Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
Gross D J and Mezard 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 Mezard M 1989 *J. Physique* **50** 3057
Mezard M, Parisi G and Virasoro M 1987 *Spin-Glass Theory and Beyond* (Singapore: World Scientific) p 37
Parisi G 1980 *J. Phys. A: Math. Gen.* **13** 1887
Personaz L, Guyon I and Dreyfus G 1985 *J. Physique Lett.* **46** L359
Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **32** 1972