Spin glass states of the anti-Hopfield model

1998 J. Phys. A: Math. Gen. 31 7447

(http://iopscience.iop.org/0305-4470/31/37/007)

J. Phys. A: Math. Gen. **31** (1998) 7447–7459. Printed in the UK PII: S0305-4470(98)90216-5

Spin glass states of the anti-Hopfield model

Kazuo Nokura†
Shonan Institute of Technology, Fujisawa 251, Japan

Received 23 December 1997

Abstract. We study the Hopfield model with an opposite interactional sign by using a one-step replica symmetry breaking ansatz and the marginality condition. We show that this model belongs to spin glass models which have a dynamical phase transition, which is not associated with usual replica instability. Although this model is motivated by the observations about unlearning, it has various interesting aspects: correlations among interactions, a simple meaning as an optimization problem, and SK-like and non-SK-like spin glass phases depending on the number of 'memorized' patterns.

1. Introduction

Recently, there have been several studies on spin glass models which are characterized by the dynamical phase transitions which are not associated with the static instability of spin glass order parameters [1–3]. Below this transition temperature, the system has many metastable states which have higher energies than the ground state. This phenomenon is very interesting and it is desirable to study how general such transitions are, especially in spin models which have the usual two-spin interactions.

In this paper, we study the Hopfield model [4] with an opposite interactional sign, which is inspired by the studies of paramagnetic unlearning in infinite range spin glass models. We call this model the anti-Hopfield model. Although the replica theory of the Hopfield model has already been established [5], the change of interactional sign gives rise to highly nontrivial spin glass models. As we shall see, this model is beyond the usual static replica theory, which may be one of the reasons why there are few descriptions of this model in the literature. In addition, the relation between unlearning and such spin glass models will be an attractive subject in the study of neural networks.

The idea of unlearning in REM sleep was originally suggested by some biologists as a mechanism to remove spurious states in neural networks [6, 7]. This idea was studied numerically for the Hopfield model and gave some promising results to improve neural networks [8, 9]. Recently, it has been discovered that unlearning of paramagnetic spin configurations also causes highly nontrivial changes in interactions [10, 11].

The advantage of this formulation is that the analytic studies of interactional changes are feasible since they are expressed by paramagnetic spin correlation functions which can be studied by high-temperature expansion. One of the important results is that the Hopfield model evolves into the pseudo-inverse model.

The formulation by paramagnetic configurations is general enough to be applied to other spin models such as the Sherrington–Kirkpatrick (SK) model [12]. In the case of the

† E-mail address: nokura@cosmos.la.shonan-it.ac.jp

0305-4470/98/377447+13$19.50

SK model, original interactions have no correlation, but the interactions after unlearning get correlated since the change of interactions are of the Hebb-rule nature. Thus, spin glass models with correlated interactions arise by paramagnetic unlearning, in particular frustrations among interactions are changed. In addition to this academic interest, we have another reason to be interested in unlearning in random neural networks. That is, according to some biological observations [13], newborns spend a much longer time in REM sleep than adults. At this age, neural networks develop greatly, presumably in a random manner. Therefore, unlearning in random networks may simulate one aspect of the development of neural networks in early life.

Let us briefly describe the relation between the anti-Hopfield model and unlearning. The energy functions of the infinite range spin models are given by

$$H = -\tfrac{1}{2}\sum_{i\neq j} J_{ij} S_i S_j \tag{1.1}$$

where S_i are spin degrees of freedom which are assumed to be Ising type ± 1 and J_{ij} are interactions. i is a site index which takes $1, 2, \ldots, N$. N is the system size. In the previous papers [10, 11], it was shown that the interactions after paramagnetic unlearning are approximately expressed by the equation

$$J'_{ij} = (1+\theta)J_{ij} - \epsilon \sum_{k\neq i,j} J_{ik} J_{kj} \tag{1.2}$$

for small interactional changes. We can see that, after many iterations of (1.2), the Hopfield model evolves into the pseudo-inverse model for suitable values of θ and ϵ (see appendix A). When J_{ij} are SK interactions, which have no correlation among interactions, the second terms induce correlation among J'_{ij}. Here, instead of studying the evolution, we ask if there are any spin glass models which have interactional correlations suggested by (1.2). After several inspections we realize that, assuming the SK interactions for J_{ij}, the model without the first term looks very interesting. As we will see in section 2, this model is very close to the Hopfield model with an opposite interactional sign, i.e. the anti-Hopfield model.

This paper is organized as follows. In section 2 we discuss several aspects of the anti-Hopfield model, especially the eigenvalue distributions of the interaction matrix. In sections 3 and 4 replica theory for the model is discussed up to one-step replica symmetry breaking (RSB) ansatz. Section 5 is devoted to some discussions.

2. Model description

In this section, we discuss several aspects of the anti-Hopfield model, whose interactions are given by

$$J^a_{ij} = -\frac{1}{N}\sum_{\mu=1}^{P} \xi^\mu_i \xi^\mu_j \tag{2.1}$$

where ξ^μ_i are random quenched variables which take ± 1 with probability $\frac{1}{2}$, and $\mu = 1, 2, \ldots, P$. In the following, we discuss various aspects of (2.1), especially the eigenvalue distributions of the interaction matrix, which reveal the interesting relations to other models.

Let us first discuss the relation between (2.1) and (1.2) with $1+\theta = 0$ and $\epsilon = 1$ for the SK model. Then J_{ij} in (1.2) obey $\overline{J^2_{ij}} = 1/N$, where the overbar denotes a sample average, and no correlations among J_{ij}. Then the correlations among J'_{ij} are given by

$$\overline{J'_{ij} J'_{jl} \ldots J'_{li}} = (-1)^k \frac{(N-k)}{N^k} \tag{2.2}$$

where the number of J'_{ij} is k and site indices are all different.

In the thermodynamic limit, this expression becomes the same as the anti-Hopfield model with the number of patterns N. In the mean-field theories such as replica method, only correlations of interactions of the type (2.2) appear as discussed in appendix B. Thus (1.2) without the first term are well approximated by the anti-Hopfield interactions with $\alpha \equiv P/N = 1$.

Another interesting aspect of the anti-Hopfield model lies in the fact that it is an optimization problem whose cost function has a simple meaning. With (2.1), the energy function becomes

$$H = \frac{1}{2N}\sum_{\mu}\left(\sum_{i}\xi_i^{\mu}S_i\right)^2 - \frac{1}{2}N\alpha. \tag{2.3}$$

In this form, we notice that the problem becomes a search for spin configurations which are as orthogonal as possible to all ξ^{μ}.

For $P \sim 1$, which includes the anti-ferromagnet, there are many solutions which satisfy the condition. The cases $P \sim N$ show very interesting properties depending on P, which will be presented in this paper.

Now we discuss the eigenvalue distribution of (2.1). It is commonly believed that the eigenvectors with the largest eigenvalue of the interactional matrix govern the nature of low-temperature spin glass states, especially in the studies of the TAP equation [14]. This point of view is significantly helpful to discuss the relation between the anti-Hopfield model and other spin glass models.

The interaction matrices are generally expressed by the form

$$J_{ij} = \sum_{\lambda}\langle i|\lambda\rangle J_{\lambda}\langle\lambda|j\rangle \tag{2.4}$$

where $\langle i|\lambda\rangle$ are eigenvectors of J_{ij}. For the anti-Hopfield model, by setting $J_{\lambda} \to -J_{\lambda}$ in the known results [15], we obtain the eigenvalue distribution

$$\rho(J_{\lambda}) = \begin{cases} (1-\alpha)\delta(J_{\lambda}) + \rho_0(J_{\lambda}) & \alpha < 1 \\ \rho_0(J_{\lambda}) & \alpha \geqslant 1 \end{cases} \tag{2.5}$$

where

$$\rho_0(x) = \frac{(4\alpha - (x+\alpha+1)^2)^{1/2}}{2\pi|x|}. \tag{2.6}$$

In these expressions, diagonal elements $J_{ii} = -\alpha$, which are irrelevant in thermodynamics, are included for the sake of simplicity. If we want to impose $J_{ii} = 0$, J_{λ} should be replaced by $J_{\lambda} + \alpha$ in (2.4). Note that $\rho(J_{\lambda})$ for $\alpha < 1$ has two parts: $\rho_0(x)$ which is nonzero for $-(\sqrt{\alpha}+1)^2 < x < -(\sqrt{\alpha}-1)^2$ and the delta function which is located at $J_{\lambda} = 0$. Thus the largest eigenvalue is $J_{\lambda} = 0$ which is strongly degenerated. The reason for the delta function at $J_{\lambda} = 0$ is that the Hopfield interaction matrix is a projection operator onto a P-dimensional vector subspace made of ξ^{μ}. In the case of the anti-Hopfield model, the configurations perpendicular to this space make the energy function minimum, whereas the configurations on this space give higher energies. For $\alpha = 1$, $\rho(x)$ has a singularity $1/\sqrt{|J_{\lambda}|}$ at the largest eigenvalue. For $\alpha \gg 1$, $\rho(x)$ becomes similar to the semicircle law. Thus we expect that the phase transitions becomes similar to that in the SK model as α increases, whereas for $\alpha \sim 1$ or smaller, the anti-Hopfield model will resemble the random orthogonal model (ROM), which is obtained by replacing J_{λ} with ± 1 in (2.4) [16].

Another spin glass model which has this property is the pseudo-inverse model [17], which has only two eigenvalues in the interaction matrix. Studies of this model imply

that the spin glass phase is rather different from the SK model [18, 19]. Interestingly, the pseudo-inverse model arises from the Hopfield model by paramagnetic unlearning [10, 11]. We discuss this point in appendix A, which is also a supplementary note for the previous paper.

The discussion above implies that the anti-Hopfield model with $\alpha < 1$ has spin glass phases very similar to the pseudo-inverse model and ROM. On the other hand for $\alpha \gg 1$, the model will be similar to the SK model. In the next two sections, we discuss the replica theory of the anti-Hopfield model to study these speculations.

3. Replica symmetry theory

This section is devoted to the description of replica symmetry (RS) theory of the anti-Hopfield model and section 4 to one-step RSB theory.

The replica theory of the Hopfield model was developed in [5]. The free energy of the anti-Hopfield model is obtained simply by changing $\beta \to -\beta$ in equation (2.7) in [5]. In appendix B, we re-derive the expression of the free energy by using correlations of interactions, which may justify the equivalence between (2.1) for $P = N$ and (1.2) for $1 + \theta = 0$ in the thermodynamic limit, as discussed in section 2. Order parameters $q_{\alpha\beta}$ and $\lambda_{\alpha\beta}$ are also defined in appendix B.

Let us first discuss the RS solution defined by $q_{\alpha\beta} = q$ and $\lambda_{\alpha\beta} = \lambda$. Then the free energy density f is given by

$$\beta f = \frac{\alpha}{2}\left\{\ln(1+\beta(1-q)) + \frac{\beta q}{1+\beta(1-q)} - \beta\right\} - \frac{1}{2}\lambda q + \frac{1}{2}\lambda - \int \mathrm{D}z\ \ln 2\cosh(\sqrt{\lambda}z) \tag{3.1}$$

which yields the saddle point equations

$$q = \int \tanh^2(\sqrt{\lambda}z)\,\mathrm{D}z$$
$$\lambda = \frac{\alpha\beta^2 q}{(1+\beta(1-q))^2}$$

where $\mathrm{D}z = \exp(-z^2/2)\,\mathrm{d}z/\sqrt{2\pi}$. The continuous phase transition point is given by $T_{\mathrm{RS}} = -1 + \sqrt{\alpha}$, which is negative for $\alpha < 1$. Therefore the SK-like spin glass transition does not occur at least for $\alpha < 1$. Instead of searching for the discontinuous RS transition point, we will study more general one-step RSB solutions in the next section.

High-temperature expressions for f, entropy s, and energy e are given by

$$f = \frac{\alpha}{2\beta}(\ln(1+\beta) - \beta) - \frac{1}{\beta}\ln 2$$
$$s = \frac{\alpha}{2}\left\{\frac{\beta}{1+\beta} - \ln(1+\beta)\right\} + \ln 2 \tag{3.2}$$
$$e = -\frac{\alpha}{2}\frac{\beta}{1+\beta}.$$

Note that s becomes negative for $T < T_s$, which is defined by $s = 0$. T_s is 0.113 for $\alpha = 1$, and $\sim \exp(-1 - 2\ln 2/\alpha)$ for small α.

Figure 1 shows T_{RS} and T_s as functions of α, as well as T_g which will be introduced in section 4. For $\alpha < 1.4$, the entropy becomes negative before RS instability take places. This situation has been found in various complex models such as Ising perceptron and p-spin models. For $\alpha > 1.4$, $T_{\mathrm{RS}} > T_s$, which is similar to the SK model. Thus the properties

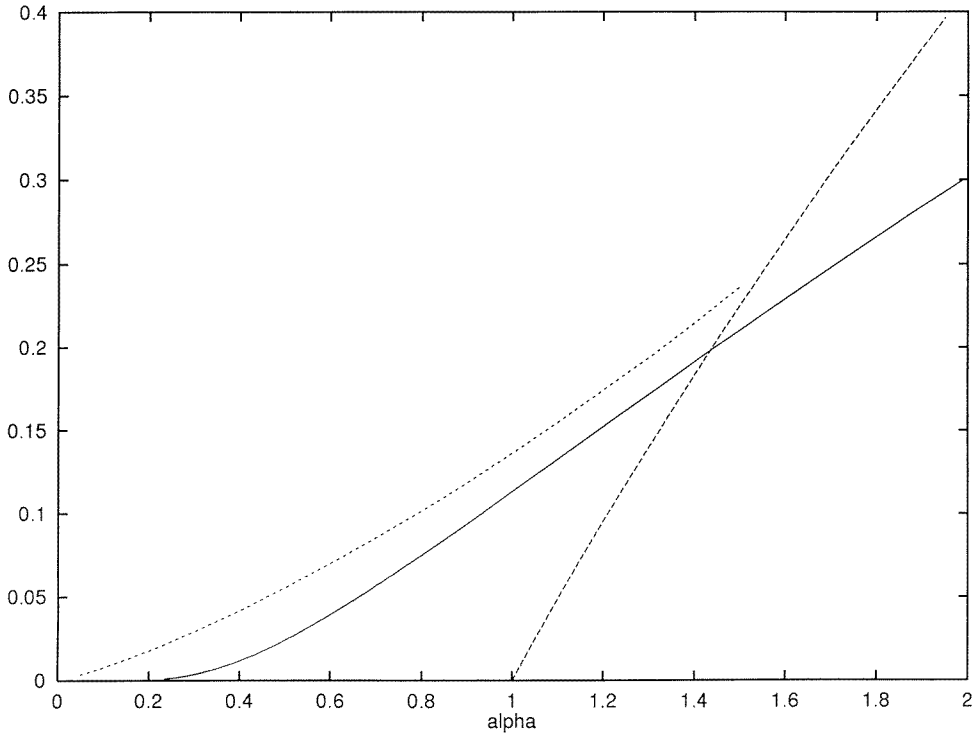

Figure 1. α-dependence of T_{RS}, T_s, and T_g from bottom. $T_s > T_{\mathrm{RS}}$ for $\alpha < 1.4$. Spin glass transition temperature T_g, which is defined in section 4, is also shown for $\alpha < 1.5$.

of the anti-Hopfield model will be quite different depending on the value of α. This is consistent with the discussion about the eigenvalue distributions in section 2.

We have some remarks on these results. First, the expression (3.2) is very similar to the Golay–Bernasconi approximation of the low-autocorrelation problem [20]. This is due to the fact that both problems involve finding spin configurations which satisfy many constraints. Actually, at high temperatures, each $m_\mu = \sum_i \xi_i^\mu S_i/\sqrt{N}$ may obey the independent Gaussian distribution, which leads to (3.2). Secondly, we can get an idea about the minimum of the cost function (2.3) by assuming that it is achieved at T_s. With this assumption, we obtain $e_{\mathrm{min}} = -\alpha/2 + \alpha/2\exp(-1 - 2\ln 2/\alpha)$ for small α. Note that the second term is exponentially small for $\alpha \to 0$.

4. One-step RSB solutions

This section is devoted to the descriptions of one-step RSB solutions. The studies in section 3 imply that our model does not belong to SK-like spin glass models for $\alpha < 1.4$. In this case, we need to discuss the dynamical phase transitions to explain the results of numerical simulations, as was done for ROM. In the framework of one-step RSB, the conjecture in [3] is to impose the marginality condition, which is expected to characterize the dynamical phase transition. In the following, we shall report the results of both one-step RSB solutions as well as the Monte Carlo (MC) simulations.

4.1. Static one-step RSB

In the one-step RSB ansatz, the order parameter matrix $q_{\alpha\beta}$ is divided into $(n/m) \times (n/m)$ sub-blocks of the size $m \times m$. Assuming $q_{\alpha\beta} = q_1$, $\lambda_{\alpha\beta} = \lambda_1$ if α and β belong to the same diagonal sub-blocks, and $q_{\alpha\beta} = q_0$, $\lambda_{\alpha\beta} = \lambda_0$ if not, the free energy density reduces to

$$\begin{aligned}\beta f = \frac{\alpha}{2}\left\{\frac{1}{m}\ln(1+\beta x_m) + \left(1-\frac{1}{m}\right)\ln(1+\beta x_0) + q_0\frac{\beta}{1+\beta x_m} - \beta\right\} \\ +\tfrac{1}{2}\{(m-1)\lambda_1 q_1 - m\lambda_0 q_0\} + \tfrac{1}{2}\lambda_1 \\ -\frac{1}{m}\int \mathrm{D}z \,\ln\int \mathrm{D}y\,\left\{2\cosh\left(\sqrt{\lambda_0}z + \sqrt{\lambda_1-\lambda_0}y\right)\right\}^m \end{aligned} \tag{4.1}$$

where $x_m = 1 + m(q_1 - q_0) - q_1$ and $x_0 = 1 - q_1$. The details of the derivation are given in appendix B.

Let us first discuss the results for $\alpha = 1.0$, which corresponds to unlearning in the SK model. Let T_1 be the temperature at which static one-step RSB solutions appear, which satisfy $\partial f/\partial\lambda_0 = \partial f/\partial\lambda_1 = 0$, $\partial f/\partial q_0 = \partial f/\partial q_1 = 0$, $\partial f/\partial m = 0$ with $0 < m < 1$. Numerical studies of the saddle point equations reveal that q_1 starts from about 0.93 and m starts from 1.0 at $T = T_1 = 0.117$, whereas q_0 and λ_0 are practically zero for all temperatures. Note that T_1 is just above $T_s = 0.113$. Numerical studies reveal that T_1 tends to T_s as α decreases. The energy e obtained from these solutions is presented in figure 2, which takes about -0.448 and does not vary much in the low-temperature phase. In this figure, the results of MC simulation and marginally stable solution (see below) are also shown.

For the RSB ansatz, there are many studies on the stability of the solutions ([27] and references therein). The most interesting modes are so-called replicon modes, which are related to further steps RSB. In appendix C, we review the derivation of the eigenvalues for our case, which is given by $1 - g\mu$, where

$$g = -\frac{\alpha\beta^2}{(1+\beta x_0)^2} \tag{4.2}$$

$$\mu = -\frac{\int \mathrm{D}z\,\{\cosh(\sqrt{\lambda_1}z)\}^m \cosh^{-4}(\sqrt{\lambda_1}z)}{\int \mathrm{D}z\,\{\cosh(\sqrt{\lambda_1}z)\}^m}. \tag{4.3}$$

In the one-step RSB of the SK model, $1 - g\mu$ is negative, signalling further-steps RSB. For our model with $\alpha < 1.4$, it is found to be positive in one-step RSB solutions near T_1, but as T decreases, it becomes negative.

The numerical estimates for e are obtained by MC simulations by decreasing the temperature step by step. In the figures, we present the results which are obtained by averaging over 2×10^4 MC steps at each temperature. The results of MC simulations show a change of the temperature dependence of e at a temperature slightly higher than T_1. The numerical estimate of e is about -0.440 in the wide range of T, which is also higher than the value of one-step RSB. As we shall see, these discrepancies are greatly enhanced for smaller α. It is not possible to get better agreements by introducing higher-step RSB since replicon modes have positive eigenvalues at least near T_1 in these one-step RSB solutions.

4.2. One-step RSB with the marginality condition

Having the discrepancy between MC simulations and usual RSB, we study the conjecture described in [3], which is to impose the marginality condition to determine m. This condition is originally introduced as a dynamical property of spin glass models [21, 22]. It is believed

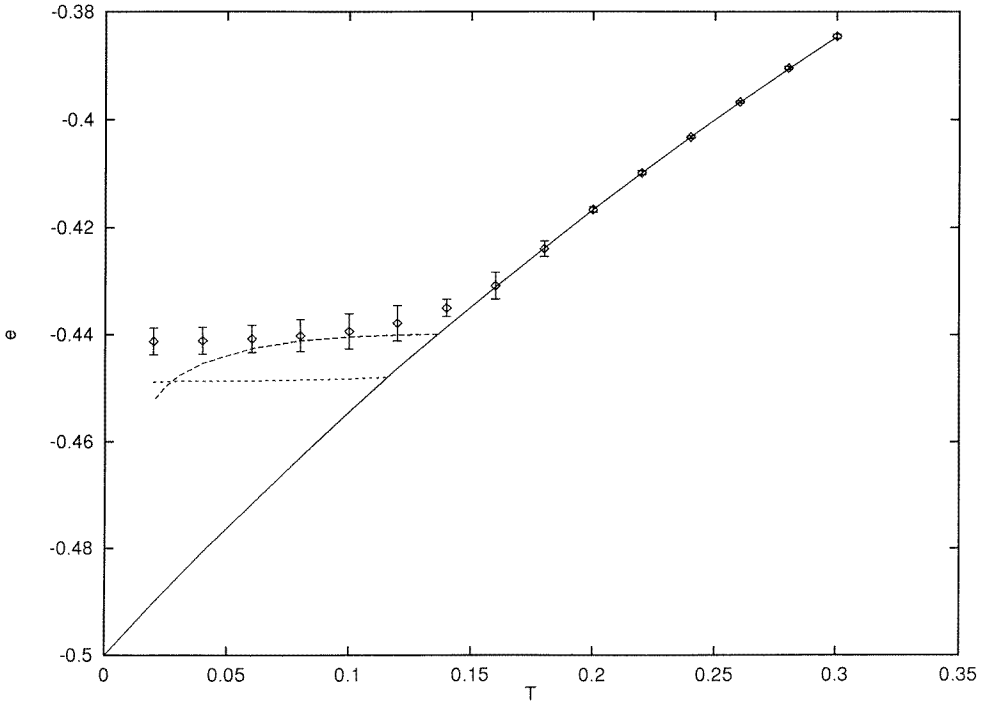

Figure 2. T-dependence of e obtained by several methods for $\alpha = 1.0$ with $N = 200$. The full curve is RS, the broken curve is one-step RSB with the marginality condition, and the dotted curve is static one-step RSB. The results of MC simulations are denoted by dots with error bars, each of which is obtained by averaging over the 10 samples. For each sample, e is obtained by averaging over 2×10^4 MC steps at each temperature.

that the dynamical stability is marginal in the glassy phase because of the non-exponential decay of autocorrelation functions. In the SK model, this situation may be achieved by many steps RSB. As discussed above, static RSB for the anti-Hopfield model is stable at least near the transition temperature. According to [3], we should search the replica solution with the marginality condition to describe the dynamical phase transition of the MC simulations. The following studies provide another test of this conjecture.

Using g and μ, the marginality condition for one-step RSB reads

$$1 - g\mu = 0. \tag{4.4}$$

To obtain the solution with this condition, $\partial f/\partial m = 0$ in the saddle point equations is replaced by (4.4) in numerical calculation. Let T_g be the temperature at which the marginal one-step RSB solution appears. Numerical studies reveal that the solutions with the marginality condition appear at $T = T_g = 0.137$ for $\alpha = 1$ and gives $e = -0.440$ in the wide range of $T < T_g$. As shown in figure 2, T_g and e seem to be in good agreement with the results obtained by MC simulations. However, for lower T, e starts to decrease and become smaller than the energies obtained by usual one-step RSB equations. At the temperature where these two energies meet, two solutions coincide. These results imply that the one-step RSB with the marginality condition is in better agreement with the results obtained by simulation at least near T_g. However for smaller T, it fails to agree with the simulation results.

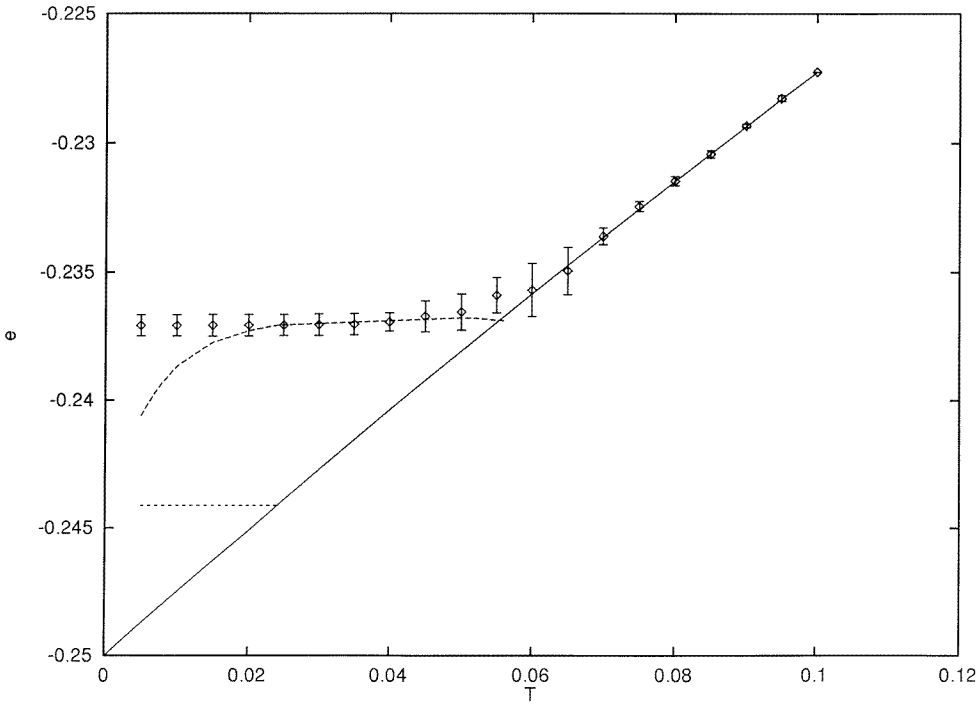

Figure 3. T-dependence of e obtained by several methods for $\alpha = 0.5$ with $N = 200$. The full curve is RS, the broken curve is one-step RSB with the marginality condition, and the dotted curve is static one-step RSB. The results of MC simulations are denoted by dots with error bars, each of which is obtained by averaging over the 10 samples. For each sample, e is obtained by averaging over 2×10^4 MC steps at each temperature.

These studies are easily generalized to other α. In the following, we shall describe the results for $\alpha = 0.5$ and 4.0.

In figure 3, the results for $\alpha = 0.5$ are presented. In this case marginal RSB gives $T_g = 0.056$ and q_1 starts from 0.98 with $m = 1.0$. e takes -0.237 for moderate $T < T_g$. On the other hand static RSB gives $T_1 = 0.024$ and q_1 starts from 0.995 with $m = 1.0$, and $e = -0.244$ in the wide range of the low-temperature region. The difference between the two kinds of one-step RSB solutions is much larger than the case $\alpha = 1.0$. Note that T_g is two times larger than T_1 as well as $e + \alpha/2$ for $\alpha = 0.5$. As shown in figure 3, the results of MC simulations are in good agreement with the marginal one-step RSB solutions like the case $\alpha = 1.0$. For small α, T_1 is located just above T_s, where RS entropy becomes zero. This means that practically $T_s \sim T_1$ and the entropy of static RSB is very close to zero. After studying several small α, we noticed that q_1 at T_g tends to 1 as α decreases. These situations are very similar to the Ising perceptron [23] and p-spin model [24].

Figure 4 shows the results for $\alpha = 4.0$, where the theoretical energy was obtained by the static one-step RSB solution. In this case, static RSB solutions appear at $T_{\mathrm{RS}} = 1$, where q_1 and q_0 appear continuously from zero. m starts from 0, taking the maximum value 0.53 near $T = 0.65$ and tends to zero as T decreases. Numerical e are slightly higher than the value of one-step RSB, which will be improved by taking into account higher-steps RSB since $1 - g\mu$ is negative for these solutions. These aspects are very similar to the SK model, as expected from the discussions about the interactional eigenvalues.

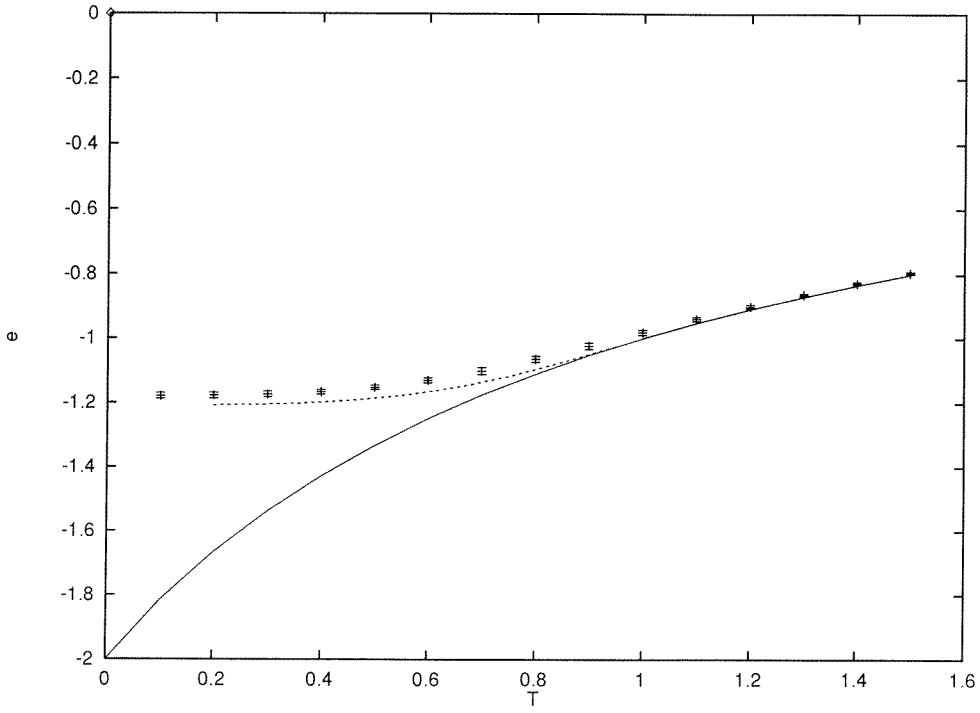

Figure 4. T-dependence of e obtained by several methods for $\alpha = 4.0$ with $N = 200$. In this case, only static one-step RSB results are presented at low temperatures.

5. Discussions

In this paper, we have studied the anti-Hopfield model in the framework of one-step RSB with the marginality condition. The model shows the behaviour which varies from SK-like to ROM-like by changing α. For $\alpha = 0.5$ and 1.0, one-step RSB solutions with the marginality condition show good agreement with MC simulations at least near T_g. Internal energies obtained by MC simulations are practically constant at low temperatures, whereas the energy of marginal one-step RSB decreases as T decreases. The marginal RSB theory with a reasonable $T \to 0$ limit remains to be studied. For $\alpha = 4.0$, the usual RSB shows good agreement with the simulation results. These phenomena, which look rather strange at first sight, are reasonable because of the drastic change of the interactional eigenvalue distribution around $\alpha \sim 1$, as discussed in section 2. For $\alpha < 1$, the largest eigenvalue degenerates strongly, which may induce numerous states of similar energies. This may be the reason why the dynamical phase transitions occur. The crossover around $\alpha \sim 1.4$ may be an interesting subject to study, although the numerical studies of the saddle point equations will demand much computer time.

Also in section 2, we have discussed that the anti-Hopfield model can be viewed as an optimization problem. This point of view can be generalized in the following way. In some optimization problems [25], the cost functions take the form

$$H = \tfrac{1}{2} \sum_{\mu=1}^{P} (f_\mu)^2 \tag{5.1}$$

where f_μ are some functions of N spin variables S_i. If f_μ are linear functions, H becomes an energy function of two-spin models, probably with external fields. Usually, the case $P \sim N$ will be the most interesting situation. There have been some attempts to solve this kind of problem by simulated annealing. Our studies imply that it works well to achieve the low-cost states only if $P \gg N$, whereas the case $P \sim N$ seems rather difficult because of the existence of the dynamical phase transition. This point was discussed for the Ising perceptron problem [22]. In our model, the relative location of $T_{\rm RS}$ and T_s depends on the parameter P. It is interesting to find that the minimum of H becomes easier as P increases, although the minimum does not imply $f_\mu \sim 0$ for all μ when P is large. As another parameter, the effect of external fields in the anti-Hopfield model will be also an interesting subject.

The anti-Hopfield model was originally motivated by the studies of unlearning in spin glass models. Learning or unlearning in spin glass models induces correlations among interactions. In the case of unlearning, frustrations among interactions increase. We believe that the anti-Hopfield model reflects this aspect. As discussed in this paper, the spin glass phase of this model for small α is similar to ROM rather than the SK model. Although we need to do further studies, we expect that the energy landscapes get more rugged and many fixed points of spin dynamics are created by unlearning. This may imply that remnant properties increase, which might work as some kind of very short-term memory in the context of neural networks. This point deserves further investigation.

Appendix A

This appendix is devoted to the discussion about unlearning in the Hopfield model and the pseudo-inverse interactions. It is known that the pseudo-inverse interactions have only two eigenvalues. We discuss it from a different point of view.

Let us study the evolution of J_{ij} for the Hopfield model with $\alpha < 1$. Assuming J_λ depend on time, (1.2) and (2.4) gives the evolution equation for eigenvalues.

$$\frac{{\rm d}J_\lambda(t)}{{\rm d}t} = cJ_\lambda(t) - J_\lambda^2(t) \tag{A.1}$$

where c is a positive constant. When the Hopfield interactions

$$J_{ij}(0) = \frac{1}{N}\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \tag{A.2}$$

are initial conditions including $J_{ii} = \alpha$, the initial eigenvalue distribution is obtained by $J_\lambda \to -J_\lambda$ in (2.5), which means all eigenvalues are positive except $J_\lambda = 0$. The equation (A.1) implies that all positive $J_\lambda(0)$ tend to c, whereas $J_\lambda = 0$ does not change. This yields

$$J_{ij}(\infty) = c\sum_{J_\lambda > 0} \langle i|\lambda\rangle\langle\lambda|j\rangle \tag{A.3}$$

where $\langle i|\lambda\rangle$ are eigenvectors of (A.2). Note that $J_{ij}(\infty)$ has two eigenvalues 0 and c.

Now we show that $J_{ij}(\infty)$ are the pseudo-inverse interactions made of ξ_i^μ. Since $\langle i|\lambda\rangle$ with $J_\lambda > 0$ and ξ_i^μ make the same vector subspace, we can set

$$\xi_i^\mu = \sqrt{N}\sum_\lambda a_{\lambda\mu}\langle i|\lambda\rangle. \tag{A.4}$$

This yields $J_\lambda = \sum_\mu a_{\lambda\mu}^2$, which implies that $a_{\lambda\mu}$ is really $P \times P$ matrix because $J_\lambda = 0$ for $N - P$ eigenvectors. The pattern correlation matrix is given by

$$\begin{aligned} C_{\mu\nu} &\equiv \sum_i \xi_i^\mu \xi_i^\nu / N \\ &= \sum_\lambda a_{\lambda\mu} a_{\lambda\nu}. \end{aligned}$$

This and (A.4) give

$$J_{ij}(\infty) = \frac{c}{N} \sum_{\mu\nu} \xi_i^\mu C^{-1\mu\nu} \xi_j^\nu \tag{A.5}$$

which are the pseudo-inverse interactions.

Appendix B

In this appendix we describe the derivation of replica free energy for the Hopfield model by using the correlation functions of interactions. The free energy density f is expressed by $\overline{f} = -\overline{\ln Z}/\beta N$, where Z is a partition function. By using the replica formula, $\overline{\ln Z} = \lim_{n\to 0}(\overline{Z^n} - 1)/n$, the problem reduces to the study of the expression given by

$$Z^n = \sum_S \exp -\tfrac{1}{2}\beta \sum_{i\neq j} J_{ij}\Omega_{ij} \tag{B.1}$$

where $\Omega_{ij} = \sum_\alpha S_i^\alpha S_j^\alpha$. After expanding Z^n in terms of J_{ij}, we have

$$Z^n = \sum_S \prod_{i<j} (1 - \beta J_{ij}\Omega_{ij} + \tfrac{1}{2}\beta^2 (J_{ij}\Omega_{ij})^2 + \cdots). \tag{B.2}$$

Each J_{ij} can be represented as a line with a site at each end. After averaging over ξ_i^μ for the anti-Hopfield model, the contributions which remain are loops made of these lines. These loops can share sites and lines, but these are expressed as a product of loops which visit each site once, which are denoted by

$$J^{(k)} = \overline{J_{ik} J_{kl} \ldots J_{li}} \tag{B.3}$$

where the number of J_{ij} is k and the site indices are all different. $J^{(k)}$ are independent of the site indices because of averaging over ξ_i^μ. Exponentiating the contributions and disregarding higher order terms of $1/N$, we obtain

$$\overline{Z^n} = \sum_S \exp \sum_{k=2}^{\infty} \frac{\beta^k J^{(k)}}{2k} \operatorname{Tr} \Omega^k. \tag{B.4}$$

Although the site indices in $\operatorname{Tr}\Omega^k$ are also different, this restriction can be removed to the leading order of N. Then $\operatorname{Tr}\Omega^k = N^k \operatorname{Tr} q^k$, where q is a spin glass order parameter matrix which has elements $q_{\alpha\beta} = \sum_i S_i^\alpha S_i^\beta / N$, including $q_{\alpha\alpha} = 1$. Using $J^{(k)} = (-1)^k \alpha / N^{k-1}$ for the anti-Hopfield model, we obtain

$$\overline{Z^n} = \sum_S \exp -\frac{N}{2} \alpha \operatorname{Tr}\{\ln(1 + \beta q) - \beta q\}. \tag{B.5}$$

Now, by inserting

$$\begin{aligned} 1 &= \int \prod_{\alpha<\beta} N \, \mathrm{d}q_{\alpha\beta} \delta\left(N q_{\alpha\beta} - \sum_i S_i^\alpha S_i^\beta \right) \\ &= \int\int_{-i\infty}^{i\infty} \prod_{\alpha<\beta} \frac{N \, \mathrm{d}q_{\alpha\beta} \, \mathrm{d}\lambda_{\alpha\beta}}{2\pi \mathrm{i}} \exp -\lambda_{\alpha\beta} \left(N q_{\alpha\beta} - \sum_i S_i^\alpha S_i^\beta \right) \end{aligned}$$

we obtain

$$n\beta f = \tfrac{1}{2}\alpha \,\mathrm{Tr}\{\ln(1+\beta q) - \beta q\} + \tfrac{1}{2}\sum_{\alpha\neq\beta} q_{\alpha\beta}\lambda_{\alpha\beta} - \ln\sum_{S}\exp\tfrac{1}{2}\sum_{\alpha\neq\beta}\lambda_{\alpha\beta}S^{\alpha}S^{\beta}. \tag{B.6}$$

For the Hopfield model, $J^{(k)} = \alpha/N^{k-1}$, which leads to $\beta \to -\beta$ in the right-hand side of (B.6).

In the one-step RSB ansatz, matrix $q_{\alpha\beta}$ is divided into $n/m \times n/m$ sub-blocks of the size $m \times m$, where diagonal sub-blocks have elements q_1 and off-diagonal sub-blocks have elements q_0. For $\lambda_{\alpha\beta}$, we assume λ_1 and λ_0 in the same manner. Then q has eigenvalues, $1 + m(q_1 - q_0) - q_1 + nq_0$, $1 + m(q_1 - q_0) - q_1$, and $1 - q_1$ with degeneracies 1, $n/m - 1$, and $n - n/m$ respectively. This leads to (4.1) in the $n \to 0$ limit.

Appendix C

This appendix is devoted to the calculation of the eigenvalue of the replicon mode for one-step RSB. There are many references on this subject ([27] and references therein). We mainly follow [3], although the description here will be restricted to the minimum which is needed in this paper. Small changes from one-step RSB are denoted by $\delta q_{\alpha\beta}$ and $\delta\lambda_{\alpha\beta}$. Let us set

$$\begin{aligned} g(x) &= \frac{\alpha}{2}(\ln(1+x) - x) \\ &= \sum_{k=2}^{\infty} c_k x^k. \end{aligned} \tag{C.1}$$

Then the second-order terms in $n\beta f$ are given by

$$n\beta\delta^2 f = \tfrac{1}{2}\sum_{(\alpha\beta)(\gamma\delta)} G_{(\alpha\beta)(\gamma\delta)}\delta q_{\alpha\beta}\delta q_{\gamma\delta} + \sum_{(\alpha\beta)}\delta q_{\alpha\beta}\delta\lambda_{\alpha\beta} + \tfrac{1}{2}\sum_{(\alpha\beta)(\gamma\delta)} F_{(\alpha\beta)(\gamma\delta)}\delta\lambda_{\alpha\beta}\delta\lambda_{\gamma\delta} \tag{C.2}$$

where

$$\begin{aligned} G_{(\alpha\beta)(\gamma\delta)} &= \frac{\partial^2 \,\mathrm{Tr}\, g(\beta q)}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}} \\ &= \sum c_k \beta^k 2k \frac{\partial (q^{k-1})_{\alpha\beta}}{\partial q_{\gamma\delta}} \end{aligned} \tag{C.3}$$

$$F_{(\alpha\beta)(\gamma\delta)} = \langle S_\alpha S_\beta\rangle\langle S_\gamma S_\delta\rangle - \langle S_\alpha S_\beta S_\gamma S_\delta\rangle. \tag{C.4}$$

In one-step RSB with $q_0 = \lambda_0 = 0$, the nonzero elements of (C.3) and (C.4) take six different values respectively. Three of them which have replica indices in the same sub-blocks contribute to the replicon mode. Let us concentrate on this case. For $G_{(\alpha\beta)(\gamma\delta)}$, they are denoted by P' for $(\alpha\beta) = (\gamma\delta)$, Q' for $\alpha = \gamma$, $\beta \neq \delta$, and R' for different α, β, γ, δ. For $F_{(\alpha\beta)(\gamma\delta)}$, P, Q and R, are expressed by

$$\begin{aligned} P &= I_2^2 - 1 \\ Q &= I_2^2 - I_2 \\ R &= I_2^2 - I_4 \end{aligned}$$

where

$$I_L = \frac{\int \mathrm{D}z\,\{\cosh(\sqrt{\lambda_1}z\}^m \tanh^L(\sqrt{\lambda_1}z)}{\int \mathrm{D}z\,\{\cosh(\sqrt{\lambda_1}z\}^m}. \tag{C.5}$$

The diagonalizations of $G_{(\alpha\beta)(\gamma\delta)}$ and $F_{(\alpha\beta)(\gamma\delta)}$ are the same as the studies of RS instability [26]. Then, in the space of a replicon mode, the matrix reduces to

$$\begin{pmatrix} g & 1 \\ 1 & \mu \end{pmatrix}$$

where $g = P' - 2Q' + R'$ and $\mu = P - 2Q + R$. Since $\delta\lambda_{\alpha\beta}$ are actually pure imaginary, the condition that the product of two eigenvalues is positive leads to

$$1 - g\mu > 0. \tag{C.6}$$

μ is explicitly given by (4.3). To evaluate g, we need some extra calculations for P', Q', and R', which consist of the expression

$$\frac{\partial (q^{k-1})_{\alpha\beta}}{\partial q_{\gamma\delta}} = \sum_{l=0}^{k-2} \{(q^{k-2-l})_{\alpha\gamma}(q^l)_{\delta\beta} + (q^{k-2-l})_{\alpha\delta}(q^l)_{\gamma\beta}\}. \tag{C.7}$$

This expression takes three different values depending on the combination of replica indices. Each value reduces to different combinations of matrix elements of q^l. The elements of q^l are given by $(q^l)_{\alpha\alpha} = va^l + wb^l$, and $(q^l)_{\alpha\beta} = ra^l + sb^l$ for sub-blocks, where $v = r = 1/m$, $w = 1 - 1/m$, $s = -1/m$, $a = 1 + (m-1)q_1$, $b = 1 - q_1$. Using the expressions for q^l, (C.1), and (C.3), we obtain

$$\begin{aligned} g &= (v-r)^2 2\beta^2 g''(\beta a) + (w-s)^2 2\beta^2 g''(\beta b) + 2(v-r)(w-s)2\beta \frac{g'(\beta b) - g'(\beta a)}{b-a} \\ &= 2\beta^2 g''(\beta b) \end{aligned}$$

which equals $-\alpha\beta^2/(1+\beta(1-q_1))^2$ for the anti-Hopfield model.

References

[1] Kirkpatrick T R and Thirumalai D 1987 *Phys. Rev.* B 5388
[2] Crisanti A, Horner H and Sommers H J 1993 *Z. Phys.* B **92** 257
[3] Marinari E, Parisi G and Ritort F 1994 *J. Phys. A: Math. Gen.* **27** 7647
[4] Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
[5] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30
[6] Crick F and Mitchison G 1983 *Nature* **304** 111
[7] Hopfield J J, Feinstein D I and Palmer R G 1983 *Nature* **304** 158
[8] Hemmen J L V, Ioffe L B, Kühn R and Vaas M 1990 *Physica* **163** 386
[9] Wimbauer S, Klemmer N and van Hemmen J L 1994 *Neural Networks* **7** 261
[10] Nokura K 1996 *J. Phys. A: Math. Gen.* **29** 3871
[11] Nokura K 1996 *Phys. Rev.* E **54** 5571
[12] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792
[13] Roffwarg H P, Muzio J N and Dement W C 1966 *Science* **152** 604
[14] Thouless D J, Anderson P W and Palmer R G 1977 *Phil. Mag.* **35** 593
[15] Crisanti A and Sompolinsky H 1987 *Phys. Rev.* A **36** 4922
[16] Parisi G and Potters M 1995 *J. Phys. A: Math. Gen.* **28** 5267
[17] Personnaz L, Guyon I and Deyfus G 1985 *J. Physique Lett.* **46** L-359
[18] Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
[19] Dotsenko V S, Yarunin N D and Dorotheyev E A 1991 *J. Phys. A: Math. Gen.* **24** 2419
[20] Marinari E, Parisi G and Ritort F 1994 *J. Phys. A: Math. Gen.* **27** 7615
[21] Sompolinsky H and Zippelius A 1982 *Phys. Rev.* B **25** 6860
[22] Horner H 1992 *Z. Phys.* B **86** 291
[23] Krauth W and Mézard M 1987 *J. Physique* **50** 3057
[24] Gross D J and Mézard M 1984 *Nucl. Phys.* B **240** 431
[25] Kirkpatrick S, Gelatt C D and Vecchi M P Jr 1983 *Science* **220** 671
[26] de Almeida J R L and Thouless D L 1978 *J. Phys. A: Math. Gen.* **11** 983
[27] Temesvári T, De Dominicis C and Kondor I 1994 *J. Phys. A: Math. Gen.* **27** 7569