The statistical mechanics of the Ising perceptron

1993 J. Phys. A: Math. Gen. 26 1077

(http://iopscience.iop.org/0305-4470/26/5/027)

J. Phys. A: Math. Gen. 26 (1993) 1077-1089. Printed in the UK

The statistical mechanics of the Ising perceptron

J F Fontanari[†] and R Meir[‡]

[†] Instituto de Física e Química de São Carlos, Universidade de São Paulo,
Caixa Postal 369, 13560 São Carlos SP, Brazil
[‡] Department of Electrical Engineering, Technion, Haifa 32000, Israel

Abstract. We investigate the problem of loading a single-layered perceptron of Ising (± 1) weights, focusing on the cases of linear and binary output neurons. Previous studies of this problem have been made using the canonical ensemble, which in many cases require the breaking of replica symmetry. We consider here an alternative approach, based on the microcanonical ensemble, and show that many results that were obtained previously using replica symmetry breaking within the canonical ensemble can be easily obtained using the replica-symmetric assumption within the microcanonical ensemble. Since the nature of the replica symmetry breaking in many of the models with discrete weights is similar, we believe that our results can immediately be extended to other cases such as learning a rule from examples, and utilizing discrete weights of larger synaptic depth.

1. Introduction

The replica formulation of the statistical mechanics of disordered systems has become a major research tool in the study of complex systems. In particular, the problem of learning in neural networks has been exhaustively studied within that framework since the seminal paper of Gardner (1988). The popularity of the replica method stems mainly from the elegance and simplicity of a particular formulation, the so-called replica-symmetric theory: research fields where this theory was shown to be appropriate have undergone rapid progress while fields requiring a more elaborate formulation, namely Parisi's replica symmetry breaking scheme (Parisi 1980, Mézard *et al* 1987), have been much slower to progress.

In the context of learning in single-layer feedforward neural networks, the replica-symmetric theory was successfully employed in the analysis of networks of real-valued weights (Györgi and Tishby 1989, Seung *et al* 1992, Meir and Fontanari 1992a). In fact, in the problem of learning from examples this theory gives exact results, this being attributed to the connectedness of the weight space. On the other hand, the theory fails badly in describing discrete weights networks, in the region where the learned rule is *unrealizable* (Seung et al 1992, Meir and Fontanari 1992b).

With at least one remarkable exception, namely the seminal paper of Gardner, most applications of the replica method, either for real-valued or discrete weights neural networks, have been carried out within the canonical ensemble formalism. The goal of this paper is to show that the replica-symmetric theory can indeed adequately describe discrete weight networks, provided it is formulated within the microcanonical ensemble formalism. Although we focus in this paper on the random mapping problem (Gardner 1988), our results can immediately be extended to the case of learning a rule, where in any event the effects of replica symmetry breaking are usually much weaker.

0305-4470/93/051077+13$07.50

We consider a neural network consisting of N binary input units $S_i = \pm 1$ connected to one output unit σ through a set of N Ising weights $W_i = \pm 1$. The state of the output unit is determined by the equation

$$\sigma = g\left(\frac{1}{\sqrt{N}}\,\boldsymbol{W}\cdot\boldsymbol{S}\right) \tag{1}$$

with the notation $\boldsymbol{x}\cdot\boldsymbol{y} = \sum_{i=1}^{N} x_i y_i$. In this paper we consider two forms for the transfer function, $g(x) = x$ and $g(x) = \mathrm{sign}(x)$. In the former case, the resulting neural network is termed linear Ising perceptron while in the latter case, it is termed boolean Ising perceptron.

Given $P = \alpha N$ input/output pairs $(\boldsymbol{S}^l, t^l)$, where $\boldsymbol{S}^l = (S_1^l, \ldots, S_N^l)$ is the lth input and t^l is the associated correct output, we wish to compute the minimal error, $\epsilon_{\min}$, which a perceptron makes in implementing the given input/output mapping. This quantity is related to the ground-state energy E_{GS} of a spin system governed by the Hamiltonian or energy

$$E(\boldsymbol{W},\boldsymbol{S},t) = \frac{1}{c}\sum_{l=1}^{P}\left[\sigma^l(\boldsymbol{W},\boldsymbol{S}^l) - t^l\right]^2 \tag{2}$$

through $\epsilon_{\min} = E_{\mathrm{GS}}/P$. Here, $c = 2$ (4) for the linear (boolean) Ising perceptron. In what follows we consider the case where each component $S_i^l = \pm 1$ is chosen at random. For the boolean Ising perceptron $t^l = \pm 1$ is also chosen at random, independently of $\boldsymbol{S}^l$, while for the linear Ising perceptron t^l is a real number drawn from a Gaussian distribution of zero mean and unit variance. Thus, the energy itself is a random variable. In the statistical mechanics approach, the weights represent the dynamical variables (spins) and the P input/output pairs play the role of the quenched impurities.

The *storage capacity* of the network α_{c} is defined as the ratio between the maximal number of input/output pairs for which $\epsilon_{\min} = 0$ and the number of input units N. It was shown (Gardner and Derrida 1988) that for the boolean Ising perceptron the replica-symmetric theory formulated in the canonical ensemble predicts $\alpha_{\mathrm{c}}^{\mathrm{B}} = 4/\pi$ which violates the information-theoretic bound $\alpha_{\mathrm{c}}^{\mathrm{B}} \leqslant 1$. As we shall show, the situation is even worse in the case of the linear Ising perceptron, where that theory predicts a non-zero value for $\alpha_{\mathrm{c}}^{\mathrm{L}}$, while a rigorous bound asserts that $\alpha_{\mathrm{c}}^{\mathrm{L}} = 0$. In this paper we show that the microcanonical replica-symmetric theory predicts α_{c} correctly for both models and gives estimates for $\epsilon_{\min}$ which are comparable with the estimates of the canonical one-step replica symmetry breaking theory.

The remainder of the paper is organized as follows. In section 2 we present an extensive study of both the canonical and microcanonical versions of the statistical mechanics of the linear Ising perceptron, obtaining several estimates for $\epsilon_{\min}$. Section 3 is devoted to the analysis of the boolean Ising perceptron, with emphasis on the microcanonical formulation. Finally, in section 4 we summarize our results and present some concluding remarks.

2. Linear Ising perceptron

In this case (2) becomes

$$E(\boldsymbol{W},\boldsymbol{S},t) = \frac{1}{2}\sum_{l=1}^{P}\left(\frac{1}{\sqrt{N}}\boldsymbol{W}\cdot\boldsymbol{S}^l - t^l\right)^2 . \tag{3}$$

Since in the limit $N \to \infty$, the output σ^l has the same distribution as t^l, the existence of a regime at low values of α where $\epsilon_{\min}$ vanishes, thus implying a non-zero value for α_c^L, cannot be discarded *a priori*. We note that this model has been recently studied in the context of learning a rule by Seung *et al* (1992). We focus, however, on the zero-temperature limit, showing that the canonical approach to this problem leads to disastrous results with respect to the entropy.

In the following we present estimates for $\epsilon_{\min}$ obtained within the canonical and microcanonical formulations. We also derive a rigorous bound for this quantity, employing an annealed approximation to evaluate the averages over the quenched random variables.

2.1. Canonical ensemble

For a fixed realization of the input/output mapping, the neural network is assumed to be in contact with a heat bath or any source of noise characterized by the parameter β, which plays the role of the inverse temperature, $\beta = 1/T$. The probability distribution on the space of networks with average energy $E(\boldsymbol{W}, \boldsymbol{S}, t)$ is then given by the Gibbs distribution

$$P(\boldsymbol{W}) = \frac{\mathrm{e}^{-\beta E(\boldsymbol{W},\boldsymbol{S},t)}}{Z} \tag{4}$$

where the partition function Z is defined by

$$Z = \mathrm{Tr}_{\boldsymbol{W}}\ \mathrm{e}^{-\beta E(\boldsymbol{W},\boldsymbol{S},t)} \tag{5}$$

and $\mathrm{Tr}_{\boldsymbol{W}}$ *stands for the summation over the* 2^N *possible weight configurations. Since* it is believed that in the thermodynamic limit, $N \to \infty$, the self-averaging quantities are the extensive ones, we follow the standard convention (see Binder and Young 1986 for a review) and evaluate the average free energy density

$$-\beta f(\beta, \alpha) = \frac{1}{N} \langle\langle \ln Z \rangle\rangle \tag{6}$$

where $\langle\langle \cdots \rangle\rangle$ represents an average over the distributions of $\boldsymbol{S}^l$ and t^l. The replica method is a prescription for evaluating the average in (6): employing the identity

$$\langle\langle \ln Z \rangle\rangle = \lim_{n\to 0} \frac{1}{n} \ln \langle\langle Z^n \rangle\rangle \tag{7}$$

one first calculates $\langle\langle Z^n \rangle\rangle$ for integer n and then analytically continues to real $n \sim 0$.

In this formulation, the ground-state properties are obtained in the limit $\beta \to \infty$. For instance, the minimal fraction of errors is given by

$$\epsilon_{\min} = \lim_{\beta\to\infty} \frac{1}{\alpha} \frac{\partial(\beta f)}{\partial \beta} \tag{8}$$

and the ground-state entropy density is

$$s = \lim_{\beta\to\infty} \beta^2 \frac{\partial f}{\partial \beta}\,. \tag{9}$$

The evaluation of f in the limit $N \to \infty$ is standard by now (Gardner and Derrida 1988), so we present the final result only:

$$-\beta f = \lim_{n\to 0} \frac{1}{n} \operatorname{extr}\left\{ -\sum_{a<b} q_{ab}\hat{q}_{ab} + G_0(\hat{q}_{ab}) + \alpha G_1(q_{ab}) \right\} \tag{10}$$

where

$$G_0 = \ln\left\{ \prod_{a=1}^{n} \mathrm{Tr}_{W^a} \exp\left(\sum_{a<b} \hat{q}_{ab} W^a W^b \right) \right\} \tag{11}$$

$$G_1 = \ln \int \prod_a \frac{\mathrm{d}x_a}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\sum_a (1+2\beta)x_a^2 - \sum_{a<b} \beta(1+q_{ab})x_a x_b \right\} \tag{12}$$

with $\mathrm{Tr}_{W^a} = \sum_{W^a=\pm 1}$.

The free energy must be extremized with respect to the order parameters q_{ab} and $\hat{q}_{ab}$. The former order parameter has a simple physical interpretation,

$$q_{ab} = N^{-1} \langle\!\langle \langle \boldsymbol{W}_a \rangle_{\mathrm{T}} \cdot \langle \boldsymbol{W}_b \rangle_{\mathrm{T}} \rangle\!\rangle \tag{13}$$

i.e. it measures the average overlap between two equilibrium states, $\boldsymbol{W}_a$ and $\boldsymbol{W}_b$. Here, $\langle \cdots \rangle_{\mathrm{T}}$ stands for a thermal average. To proceed further, next we consider two standard ansätze for the structure of the order parameters.

2.1.1. Replica-symmetric theory. Since the replicas have no *a priori* physical meaning, it is natural to assume that the order parameters are symmetric under permutations of the replica indices, i.e.

$$q_{ab} = q \qquad \text{and} \qquad \hat{q}_{ab} = \hat{q} \qquad \forall a < b\,. \tag{14}$$

With this ansatz the evaluation of (10) is straightforward, resulting in the free energy

$$-\beta f_{\mathrm{RS}} = -\frac{1}{2}(1-q)\hat{q} + \int \mathrm{D}z \ln 2\cosh(z\sqrt{\hat{q}}) - \frac{\alpha}{2}\left[\ln(1+\beta(1-q)) + \frac{\beta(1+q)}{1+\beta(1-q)} \right] \tag{15}$$

where the parameters q and $\hat{q}$ are determined through the saddle-point equations $\partial f_{\mathrm{RS}}/\partial q = \partial f_{\mathrm{RS}}/\partial \hat{q} = 0$ and $\mathrm{D}z = \mathrm{d}z\ \exp(-z^2/2)/\sqrt{2\pi}$.

Using (8) and (9) and taking the limit $\beta \to \infty$ so that $x \equiv \beta(1-q)$ is finite we find

$$\epsilon_{\min}^{\mathrm{RS}} = \frac{1}{(1+x)^2} \tag{16}$$

$$s_{\mathrm{RS}} = -\frac{\alpha}{2}\left[\ln(1+x) - \frac{x}{1+x} \right] \tag{17}$$

where

$$x = \frac{1}{\sqrt{\alpha\pi} - 1}\,. \tag{18}$$

This solution, which possesses $\epsilon_{\min}^{\mathrm{RS}} > 0$, exists for $\alpha > 1/\pi$ only. Its entropy is negative, increasing from $-\infty$ to 0 as α increases from its lower limit to ∞.

For $\alpha < 1/\pi$, there exists a solution with $q < 1$ for which $\epsilon_{\min}^{\mathrm{RS}} = 0$ and $s_{\mathrm{RS}} = -\infty$ independently of α. Thus the replica-symmetric theory predicts $\alpha_c^{\mathrm{L}} = 1/\pi$. However, the pathologies in the entropy indicate that this theory is totally inadequate to describe the thermodynamic behaviour of the linear Ising perceptron.

2.1.2. Replica symmetry breaking. Following Parisi's scheme (Parisi 1980, Mézard *et al* 1987), we perform the first stage of replica symmetry breaking by dividing the n replicas into n/m groups of m replicas and setting $q_{ab} = q_1$, $\hat{q}_{ab} = \hat{q}_1$ if a and b belong to the same group and $q_{ab} = q_0$, $\hat{q}_{ab} = \hat{q}_0$ otherwise. The physical meaning of the variables q_0, q_1 and m is given by (Krauth and Mézard 1989)

$$q_0 = N^{-1} \langle\!\langle \langle \boldsymbol{W}_a\rangle_{\mathrm{T}} \cdot \langle \boldsymbol{W}_b\rangle_{\mathrm{T}} \rangle\!\rangle \quad (a \neq b) \qquad q_1 = N^{-1} \langle\!\langle \langle \boldsymbol{W}_a\rangle_{\mathrm{T}}^2 \rangle\!\rangle$$
$$m = 1 - \sum_a P_a^2 \,. \tag{19}$$

Physically, q_0 is the overlap between a pair of different equilibrium states, q_1 represents the self-overlap of an equilibrium state with itself, and m is the probability of finding two copies of the system in two different states (P_a is just the Gibbs probability measure for the state a). Note that in the limit $n \to 0$, the parameter m is constrained to the range $0 \leqslant m \leqslant 1$. Thus the free energy, (10), becomes

$$\begin{aligned}
-\beta f_{\mathrm{RSB}}^{(1)} = {} & \frac{1}{2}\left[m q_0 \hat{q}_0 + ((1-m)q_1 - 1)\,\hat{q}_1\right] + \frac{\alpha(1-m)}{2m} \ln(1 + \beta(1-q_1)) \\
& + \frac{1}{m}\int \mathrm{D}z_0 \ln\left\{\int \mathrm{D}z_1 \left[2\cosh\left(z_0\sqrt{\hat{q}_0} + z_1\sqrt{\hat{q}_1 - \hat{q}_0}\right)\right]^m\right\} \\
& - \frac{\alpha}{2m} \ln(1 + \beta(1-q_1) - \beta m(q_0 - q_1)) \\
& - \frac{\alpha\beta}{2} \frac{1 + q_0}{1 + \beta(1-q_1) - \beta m(q_0 - q_1)} \,.
\end{aligned} \tag{20}$$

To obtain sensible results, the limit $\beta \to \infty$ must be taken so as to keep the parameters $x \equiv \beta(1-q_1)$ and $\mathrm{D} \equiv \beta m$ finite. The motivation for making this ansatz is that it is the only one consistent with a non-zero ground-state energy (which is required by the analysis of the next section). We note that a similar assumption concerning the one-step solution was made by Crisanti *et al* (1986) when discussing replica symmetry breaking in the Hopfield (1982) model. We are left then with the task of *maximizing* $f_{\mathrm{RSB}}^{(1)}$ with respect to x, D and q_0, since the remaining order parameters, $\hat{q}_1$ and $\hat{q}_0$, can easily be eliminated in their favour (Mézard *et al* 1987). Instead of solving the saddle-point equations, we have opted for directly maximizing $f_{\mathrm{RSB}}^{(1)}$ using a standard maximization routine, namely the simplex algorithm (Nelder and Mead 1965). Our main findings are as follows. We were unable to find any solution, besides the replica-symmetric one, for $\alpha < 0.21$. However, for α larger than this value we found a solution with $\epsilon_{\min}$ larger than the corresponding replica-symmetric values. Moreover, the entropy of this solution, though still negative for all α, is larger than the replica-symmetric one. These results show a tendency towards the exact solution, i.e. one with vanishing entropy, as the number of steps of replica symmetry breaking increases. We note however that although the region where the entropy is $-\infty$ has shrunk from $\alpha \in [0, 1/\pi]$ to $\alpha \in [0, 0.21]$, it has not vanished. We expect that as the number of steps of replica symmetry breaking increases this region shrinks to zero. We should also mention that we are not aware of any other model where the entropy displays this pathological behaviour for the replica-symmetric and one-step solutions.

2.2. *Microcanonical ensemble*

The microcanonical formulation of the statistical mechanics is based on the computation of $\mathcal{N}(E)$, the number of networks with energy E. Once this basic quantity is known, the average entropy density is calculated by the Boltzmann relation

$$s(E) = \frac{1}{N} \langle\!\langle \ln \mathcal{N}(E) \rangle\!\rangle = \lim_{n\to 0} \frac{1}{Nn} \ln \langle\!\langle (\mathcal{N}(E))^n \rangle\!\rangle \tag{21}$$

from which we can obtain all the thermodynamic quantities. As before, the average indicated by $\langle\!\langle \cdots \rangle\!\rangle$ is taken over the probability distributions of $\boldsymbol{S}^l$ and t^l. Within this formulation, E_{GS}, and consequently $\epsilon_{\min}$, can be obtained by calculating the lowest value of E for which $s(E)$ is positive. Clearly, $s(E < E_{\mathrm{GS}}) \to -\infty$ since the entropy of a discrete spin system cannot be negative. The connection with the canonical formulation is made through the thermodynamic relationship

$$\partial(Ns)/\partial E = \beta\,. \tag{22}$$

We can calculate $\mathcal{N}(E)$ by introducing the quantity Y_W which is 1 if $E(\boldsymbol{W}, \boldsymbol{S}, t) = E$ and zero otherwise, so that

$$\mathcal{N}(E) = \mathrm{Tr}_W\, Y_W\,. \tag{23}$$

However, rather than $\mathcal{N}(E)$, the evaluation of (21) through the replica method requires the calculation of the nth moment $\langle\!\langle (\mathcal{N}(E))^n \rangle\!\rangle$. Noting that the energy, (3), is a random variable distributed according to the probability distribution

$$P(E') = \langle\!\langle \delta\,(E' - E(\boldsymbol{W}, \boldsymbol{S}, t)) \rangle\!\rangle \tag{24}$$

the calculation of this moment becomes straightforward:

$$\begin{aligned}\langle\!\langle (\mathcal{N}(E))^n \rangle\!\rangle &= \left\langle\!\!\left\langle \prod_{a=1}^{n} \mathrm{Tr}_{W^a}\, Y_{W^a} \right\rangle\!\!\right\rangle \\ &= \mathrm{Tr}_{W^1}\mathrm{Tr}_{W^2}\cdots\mathrm{Tr}_{W^n} P\,(E'_1 = E, E'_2 = E, \ldots, E'_n = E)\end{aligned} \tag{25}$$

where $P(E'_1 = E, E'_2 = E, \ldots, E'_n = E)$ is the joint probability that networks $\boldsymbol{W}^1, \boldsymbol{W}^2, \ldots, \boldsymbol{W}^n$ have energy E. Thus

$$\langle\!\langle (\mathcal{N}(E))^n \rangle\!\rangle = \left\langle\!\!\left\langle \prod_{a=1}^{n} \mathrm{Tr}_{W^a}\, \delta\,(E - E(\boldsymbol{W}^a, \boldsymbol{S}, t)) \right\rangle\!\!\right\rangle . \tag{26}$$

Using the integral representation of the delta function we find

$$\langle\!\langle (\mathcal{N}(E))^n \rangle\!\rangle = \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \prod_a \frac{\mathrm{d}\hat{\epsilon}_a}{2\pi\mathrm{i}} \left\langle\!\!\left\langle \prod_a \mathrm{Tr}_{W^a}\, \exp N\,[\alpha\hat{\epsilon}_a\epsilon - \hat{\epsilon}_a\epsilon(\boldsymbol{W}^a, \boldsymbol{S}, t)] \right\rangle\!\!\right\rangle \tag{27}$$

where the second term inside the average symbol is nothing but the canonical partition function replicated n times with β replaced by $\hat{\epsilon}_a$, and $\epsilon = E/\alpha N$. Keeping this correspondence in mind, it is straightforward to obtain, in the thermodynamic limit,

$$\frac{1}{N} \ln \langle\!\langle (\mathcal{N}(\epsilon))^n \rangle\!\rangle = \lim_{n\to 0} \frac{1}{n}\, \mathrm{extr}\left\{ \alpha\epsilon \sum_a \hat{\epsilon}_a - \sum_{a<b} q_{ab}\hat{q}_{ab} + \alpha G'_1(q_{ab}) + G_0(\hat{q}_{ab}) \right\} \tag{28}$$

where G_0 is given by (11), and G_1' is given by (12) with β replaced by $\hat{\epsilon}_a$. Here, the extremum is taken with respect to $\hat{\epsilon}_a$, q_{ab} and $\hat{q}_{ab}$.

Before taking the limit $n \to 0$, it is interesting to consider the case $n = 1$, the so-called annealed approximation, where s is approximated by s_A,

$$s_A(\epsilon) = \frac{1}{N} \ln \langle\langle \mathcal{N}(\epsilon) \rangle\rangle \; . \tag{29}$$

Due to the convexity of the logarithm function we have the very useful inequality

$$s_A(\epsilon) \geqslant s(\epsilon) \tag{30}$$

which implies that $s(\epsilon) \to -\infty$ for any ϵ such that $s_A(\epsilon) < 0$. This is the reason why the value of ϵ at which s_A vanishes, $\epsilon^A_{\min}$, is a rigorous lower bound to $\epsilon_{\min}$.

In the following we discuss the annealed approximation and the replica-symmetric ansatz for the order parameters which appear in (28).

2.2.1. Annealed approximation. The main appeal of this approximation is that it is free of all mathematical subtleties, while providing a rigorous lower bound to $\epsilon_{\min}$. Setting $n = 1$ in (28) and dropping the replica indices we find

$$s_A(\epsilon) = \ln 2 + \frac{\alpha}{2}(1 - \epsilon + \ln \epsilon) \tag{31}$$

where we have used the saddle-point equation $\partial s_A / \partial \hat{\epsilon} = 0$ to eliminate $\hat{\epsilon}$. Since $s_A \to -\infty$ when $\epsilon \to 0$ we conclude that $\alpha_c^L = 0$, i.e. there is no set of weights W that can perfectly implement αN random input/output associations. For small α we find

$$\epsilon^A_{\min} \approx \exp\left(-1 - \frac{2}{\alpha} \ln 2\right) \; . \tag{32}$$

It is interesting to note that for a random weight configuration ($s_A = \ln 2$) we find $\epsilon = 1$. For values larger than this, the system enters a regime where $\partial s_A / \partial \epsilon$ is negative ($\beta < 0$). Although a physical interpretation can be given to this regime (Landau and Lifshitz 1980) we do not consider this region any further, as it is not relevant to the calculation of ground-state properties. This observation is valid also within the replica framework, discussed next.

2.2.2. Replica-symmetric theory. Taking the limit $n \to 0$ in (28) with the ansatz

$$q_{ab} = q \qquad \hat{q}_{ab} = \hat{q} \quad \forall a < b \qquad \text{and} \qquad \hat{\epsilon}_a = \hat{\epsilon} \, \forall \, a \tag{33}$$

yields

$$\begin{aligned} s_{RS}(\epsilon) = {} & \alpha\epsilon\hat{\epsilon} - \frac{1}{2}(1-q)\hat{q} + \int \mathrm{D}z \;\; \ln 2\cosh(z\sqrt{\hat{q}}) \\ & - \frac{\alpha}{2}\left[\ln(1 + \hat{\epsilon}(1-q)) + \frac{\hat{\epsilon}(1+q)}{1 + \hat{\epsilon}(1-q)}\right] \end{aligned} \tag{34}$$

where the parameters $q, \hat{q}$ and $\hat{\epsilon}$ are determined through the saddle-point equations

$$q = \int \mathrm{D}t \tanh^2(t\sqrt{\hat{q}}) \tag{35}$$

$$\hat{q} = \frac{\alpha\hat{\epsilon}^2(1+q)}{(1+\hat{\epsilon}(1-q))^2} \tag{36}$$

$$\epsilon = \frac{2+\hat{\epsilon}(1-q)^2}{2(1+\hat{\epsilon}(1-q))^2}\,. \tag{37}$$

The thermodynamic relationship $\partial(Ns)/\partial E = \partial s/\partial(\alpha\epsilon) = \beta$ gives $\hat{\epsilon} = \beta$, as expected. It can be easily seen that $s_{\mathrm{RS}} \to -\infty$ when $\epsilon \to 0$ in agreement with the annealed approximation. Moreover, for $\epsilon = 1$ we find $q = \hat{q} = \hat{\epsilon} = 0$ and $s_{\mathrm{RS}} = \ln 2$. The values of ϵ at which s_{RS} vanishes give the replica-symmetric estimate for $\epsilon_{\min}$. For small α we find

$$\epsilon_{\min}^{\mathrm{RS}} \approx \exp\left(-\ln 2 - \frac{2}{\alpha}\ln 2\right)\,. \tag{38}$$

Since s_{RS} is negative and *finite* for $\epsilon < \epsilon_{\min}^{\mathrm{RS}}$, the replica-symmetric theory is clearly wrong in this region. Nevertheless, its estimate for $\epsilon_{\min}$ may be exact, provided the replica-symmetric saddle-point is locally stable and $\epsilon_{\min}^{\mathrm{RS}}$ is equal to the minimal error predicted by the full replica symmetry breaking theory. Following standard calculations (Gardner and Derrida 1988) we find that the replica-symmetric saddle-point is locally stable wherever the de Almeida–Thouless condition (de Almeida and Thouless 1978)

$$\alpha\gamma_0\gamma_1 < 1 \tag{39}$$

is satisfied. Here

$$\gamma_0 = \int \mathrm{D}t\left[1 - \tanh^2\left(t\sqrt{\hat{q}}\right)\right]^2 \qquad \gamma_1 = \left(\frac{\hat{\epsilon}}{1+\hat{\epsilon}(1-q)}\right)^2\,. \tag{40}$$

As far as the computation of $\epsilon_{\min}$ is concerned, the inequality (39) is fulfilled for $\alpha < 0.616$.

We emphasize that the theory developed in this section gives no information on the nature of the ground state, which has to do with the order parameter function $P(q)$ (Mézard *et al* 1987) and therefore can only be elucidated through a full replica symmetry breaking analysis.

2.3. *Analysis of the results*

In the previous analysis we have obtained several estimates for $\epsilon_{\min}(\alpha)$, which are summarized in figure 1. The full curve represents both the microcanonical replica-symmetric and the canonical one-step replica symmetry breaking ($\alpha > 0.21$) predictions. The results are indistinguishable within our numerical precision. The long broken curve is the canonical replica-symmetric prediction which, for $\alpha > 1/\pi$, gives a surprisingly good approximation to the more reliable results presented by the full curve. Finally, the short broken curve gives the rigorous lower bound obtained

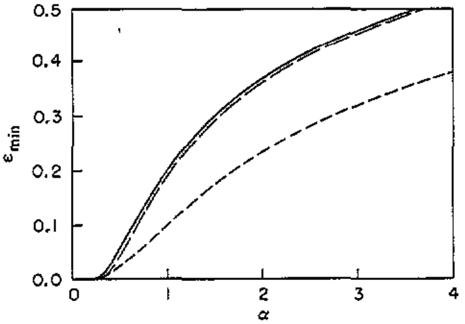

Figure 1. Estimates of $\epsilon_{\min}$ for the linear Ising perceptron as predicted by the microcanonical replica-symmetric and the canonical one-step replica-symmetry-broken theories (full curve), the canonical replica-symmetric theory (long broken curve) and the rigorous lower bound (short broken curve).

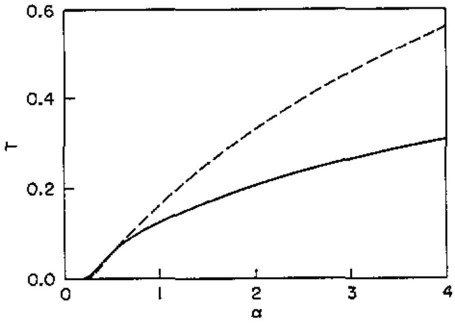

Figure 2. Phase diagram of the linear Ising perceptron in the plane (α, T) according to the canonical and microcanonical replica-symmetric theories. The full curve is the zero-entropy line and the broken curve is the Almeida–Thouless line.

within the annealed approximation. As expected, all these curves tend to 1 as α increases.

In figure 2 we present the phase diagram in the plane (α, T) obtained within the microcanonical replica-symmetric theory using the thermodynamic relationship given in (22). Obviously, the same phase diagram could be obtained using the canonical replica-symmetric theory directly. The full curve represents the zero entropy line, below which the replica-symmetric entropy is negative. The replica-symmetric saddle-point is locally stable above the broken curve (Almeida–Thouless line). These curves intersect the horizontal axis only at $\alpha = 0$. They intersect at $\alpha = 0.616$ and $T = 0.081$. For $\alpha > 0.616$ the replica-symmetry-broken solution appears as a result of the instability of the replica-symmetric solution, as in the Sherrington–Kirkpatrick (Sherrington and Kirkpatrick 1975) model. On the other hand, for $\alpha < 0.616$ one expects a first-order phase transition, in the sense of the existence of a discontinuity in the structure of the order parameter q_{ab}, though the free energy is, of course, continuous at the transition. We note that a similar phase diagram was obtained by Dotsenko and Tirozzi (1991) in their analysis of a feedback neural network with modified pseudo-inverse interactions. It would be interesting to know whether the onset of the replica symmetry breaking for $\alpha < 0.616$ occurs above the zero-entropy line, as in the model studied by Dotsenko and Tirozzi, or exactly at the zero-entropy line, as in the simplest spin glass (Gross and Mézard 1984) and the boolean Ising perceptron (Krauth and Mézard 1989). However, we do not pursue this issue any further, as it involves a rather complex numerical analysis in order to find the maxima of $f^{(1)}_{\mathrm{RSB}}$ for finite β, without being directly related to the computation of $\epsilon_{\min}$.

3. Boolean Ising perceptron

For this problem we consider the output neuron's response to be given by taking g in (1) to be the sign function. The error is then given by the number of misclassifications

$$E(W,S,t) = \sum_{l=1}^{P} \Theta\left(-t^l W \cdot S^l\right) \tag{41}$$

where the Heaviside function $\Theta(x)$ is 1 for $x > 0$ and zero otherwise.

3.1. Canonical ensemble

The problem of the boolean Ising perceptron has been considered within the canonical ensemble by Gardner and Derrida (1988) who obtained the replica-symmetric result, indicating that the correct solution requires breaking the replica symmetry. More recently, Krauth and Mézard (1989) showed how to obtain the correct solution with one level of replica symmetry breaking. In particular, it was shown that the order parameter function $P(q)$ is given by a sum of two delta functions,

$$P(q) = \frac{T}{T_c}\ \delta(q - q_0) + \left(1 - \frac{T}{T_c}\right)\ \delta(q - q_1) \tag{42}$$

where T_c is the temperature at which the replica-symmetric entropy vanishes, q_0 is the replica-symmetric overlap at T_c and $q_1 = 1$. Since the self-overlap q_1 attains its maximal value for $T \leqslant T_c$, the system is completely frozen in the low-temperature phase, being then described solely by the replica-symmetric order parameter q_0. This frozen phase is similar to the one found in the random energy model (Derrida 1981), though the energy levels of the boolean Ising perceptron are *not* independent random variables. Krauth and Mézard also found that the critical capacity for error-free learning is given by $\alpha_c^B = 0.83$, above which the ground-state energy is non-zero. Derrida *et al* (1991) have also studied this problem using toy models to obtain upper and lower bounds without recourse to the replica approach. They also pointed out the strong finite-size effects in these models.

3.2. Microcanonical ensemble

In this section we follow the basic ideas presented in section 2 for the linear perceptron. As we shall see, the minimal fraction of errors $\epsilon_{\min}$ obtained by Krauth and Mézard (1989) using replica symmetry breaking can be easily obtained within the microcanonical ensemble, using the replica-symmetric assumption. Without going into the details of the calculation, which are very similar to those given by Krauth and Mézard (1989), we find the following results. The annealed and replica-symmetric entropies are given respectively by

$$s_A(\epsilon) = (1-\alpha)\ln 2 - \alpha\left[\epsilon \ln \epsilon + (1-\epsilon)\ln(1-\epsilon)\right] \tag{43}$$

and

$$\begin{aligned} s_{RS}(\epsilon) = \alpha\epsilon\hat{\epsilon} - \tfrac{1}{2}(1-q)\hat{q} + \int \mathrm{D}z\ \ln 2\cosh(z\sqrt{\hat{q}}) \\ + \alpha \int \mathrm{D}t \ln\left[\mathrm{e}^{-\hat{\epsilon}} + (1-\mathrm{e}^{-\hat{\epsilon}})H(u)\right] \end{aligned} \tag{44}$$

where $u = t\sqrt{q/(1-q)}$, $H(u) = \int_u^\infty \mathrm{D}x$ and the variables $q, \hat{q}$ and $\hat{\epsilon}$ are determined through the saddle-point equations. It is very easy to see that the annealed results predict that for $\alpha > 1$ the minimal error is non-zero. Thus, the annealed calculation predicts $\alpha_c^B = 1$, similarly to the annealed calculation based on the canonical approach (Krauth and Mézard 1989). The condition for the local stability of the replica-symmetric saddle-point is again given by (39), with γ_0 as in (43) and γ_1 given by

$$\gamma_1 = \int \mathrm{D}t \; (\overline{x^2} - \overline{x}^2)^2 \tag{45}$$

where

$$\begin{aligned} \overline{x} &= \frac{\mathrm{i}}{\sqrt{2\pi(1-q)}} \, \frac{\mathrm{e}^{-u^2/2}}{(\mathrm{e}^{\hat{\epsilon}} - 1)^{-1} + H(u)} \\ \overline{x^2} &= -\frac{1}{\sqrt{2\pi}(1-q)} \, \frac{u\mathrm{e}^{-u^2/2}}{(\mathrm{e}^{\hat{\epsilon}} - 1)^{-1} + H(u)} \, . \end{aligned} \tag{46}$$

We have verified that the replica-symmetric saddle-point is locally stable for all α in the region $\epsilon \geqslant \epsilon_{\min}^{\mathrm{RS}}$.

3.3. *Analysis of the results*

In contrast with the linear model where $\epsilon_{\min}$ vanishes only at $\alpha = 0$, the boolean model possesses exponentially many ground states with zero error for $\alpha \leqslant 0.83$. Beyond this value, the situation seems similar to that occuring in the linear model for $\alpha > 0$, although, as the replica symmetry breaking analysis of Krauth and Mézard (1989) and the one presented in section 2 indicate, the nature of the low-temperature phase is very different.

In figure 3 we show the minimal error obtained from the microcanonical replica-symmetric approach (full curve) as well as the replica-symmetric results (long broken curve) given by Gardner and Derrida (1988). The short broken curve is the rigorous lower bound obtained using the annealed approximation. Moreover, the values of $\epsilon_{\min}(\alpha)$ predicted by the one-step solution of Krauth and Mézard are identical to the ones we obtain within the replica-symmetric microcanonical ensemble. Since the latter results are locally stable (in the sense of (39)), they lend further support to the belief of the above authors that their results are in fact exact.

The phase diagram in the (α, T) plane obtained within the replica-symmetric theory is shown in figure 4. Below the full curve, $T_c(\alpha)$, the system is frozen in its ground state while below the broken curve (Almeida–Thouless line) the replica-symmetric saddle-point is unstable. As mentioned above, the frozen phase is described by the microcanonical replica-symmetric order parameter $q_0 = q_{\min}$ which measures the overlap between two distinct states with errors equal to $\epsilon_{\min}$. In this sense, we argue that the microcanonical replica-symmetric theory can correctly describe disordered systems akin to the random energy model or the boolean Ising perceptron. However, since at the present stage of knowledge it is not possible to determine *a priori* the structure of the low-temperature phase of an arbitrary spin glass-like system, the replica symmetry breaking scheme (Parisi 1980) remains the only recourse to the understanding of such systems. Nevertheless, figure 3 attests the success of the microcanonical replica-symmetric theory in predicting correctly the storage capacity and the dependence of the minimal fraction of errors on α for the boolean Ising perceptron.

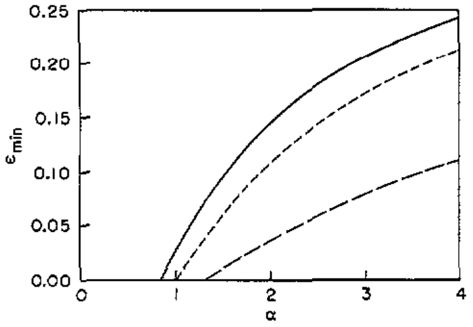

Figure 3. Estimates of $\epsilon_{\min}$ for the boolean perceptron according to the microcanonical (full curve) the canonical (long broken curve) replica-symmetric theories and the rigorous lower bound (short broken curve). The microcanonical replica-symmetric and canonical one-step replica-symmetry-breaking theories yield identical results.

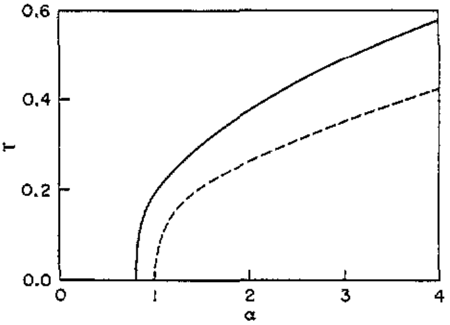

Figure 4. As figure 2, for the boolean perceptron.

4. Discussion

We have presented an analysis of the Ising perceptron both in the usual canonical, as well as in the microcanonical ensembles. We have shown that for these types of problems, the computation of the minimal fraction of errors $\epsilon_{\min}$ and the storage capacity α_c are greatly facilitated by using the microcanonical ensemble. Moreover, the annealed approximation carried out within this ensemble provides a rigorous lower bound to $\epsilon_{\min}$ and, consequently, an upper bound to α_c. We recall that in the context of the random energy model, Derrida (1981) has in fact shown that while the microcanonical calculation gives correct results, the calculation within the canonical ensemble gives wrong results.

We have focused mainly on the analysis of the linear Ising perceptron, showing that $\alpha_c^L = 0$ and presenting several estimates for $\epsilon_{\min}$. Although this system is rather uninteresting from the point of view of learning in neural networks, it can be thought of as a simple version of the integer programming problem (Garey and Johnson 1979): given a set of P pairs $(\boldsymbol{S}^l, b^l)$, is there an N-tuple $\boldsymbol{W}$ for which $\boldsymbol{W} \cdot \boldsymbol{S}^l = b^l$ for $l = 1, \ldots, P$? Here, b^l is an integer drawn from a binomial distribution which, in the limit of large N, tends to a Gaussian distribution of zero mean and variance N. We have shown that for a *typical* realization of the $P = \alpha N$ pairs $(\boldsymbol{S}^l, b^l)$ the answer to this question is 'no', provided α is non-zero. In fact, the microcanonical formulation seems to be the natural approach to study optimization problems where, usually, the goal is to estimate the average cost (or energy) of the optimal solutions. Moreover, the linear Ising perceptron seems to possess a very rich phase diagram (figure 2), since the crossing of the Almeida–Thouless and the zero-entropy lines indicates the existence of two competing replica-symmetry-broken solutions. The elucidation of this phase diagram promises to be a difficult task, as attested by the failure of the canonical approach to describe the low α, zero-temperature regime of this model.

Concerning the boolean Ising perceptron, our estimate of $\epsilon_{\min}$ lends credence to the claim of Krauth and Mézard (1989) that the one-step solution in the canonical ensemble in fact yields the exact result. Actually, we have shown that the microcanonical replica-symmetric theory can correctly describe the thermodynamics of spin-glass-like systems that possess a frozen phase akin to the one present in the random energy model (Derrida 1981). This result may be very useful, since it has been shown recently that the problem of learning unrealizable rules, i.e. rules which cannot be perfectly implemented in a given architecture, leads to behaviour patterns similar to those discussed in the context of the random mapping (Seung *et al* 1992, Meir and Fontanari 1992b). Thus, we believe it is possible to obtain the full learning curves for these problems within the microcanonical replica-symmetric ensemble. It would also be interesting to investigate whether the approach described above leads to more tractable computations in the context of multi-layered networks of binary connections (Barkai and Kanter 1991).

Acknowledgments

The research of JFF is supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and RM is supported by a Lady Davis fellowship.

References

Barkai E and Kanter I 1991 *Europhys. Lett.* **14** 107

Binder K and Young AP 1986 *Rev. Mod. Phys.* **58** 801

Crisanti A, Amit D J and Gutfreund H 1986 *Europhys. Lett.* **2** 337

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

Derrida B 1981 *Phys. Rev.* B **24** 2613

Derrida B, Griffiths R B and Prügel-Bennett A 1991 *J. Phys. A: Math. Gen.* **24** 4907

Dotsenko V S and Tirozzi B 1991 *J. Phys. A: Math. Gen.* **24** 5163

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

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

Garey M R and Johnson D S 1979 *Computers and Intractability: A Guide to the Theory of NP-Completeness* (San Francisco: Freeman)

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

Györgi G and Tishby N 1989 *Neural Networks and Spin Glasses* ed W K Theumann and R Köberle (Singapore: World Scientific)

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

Krauth W and Mézard M 1989 *J. Physique* **50** 3057

Landau L D and Lifshitz E M 1980 *Statistical Physics* (New York: Pergamon)

Meir R and Fontanari J F 1992a *Phys. Rev.* A **45** 8874

—— 1992b *J. Phys. A: Math. Gen.* **25** 1149

Mézard M, Parisi G and Virasoro MA 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)

Nelder J A and Mead R 1965 *Computer J.* **7** 308

Parisi G 1980 *J. Phys. A: Math. Gen.* **13** 1101

Sherrington S and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792

Seung S, Sompolinsky H and Tishby N 1992 *Phys. Rev.* A **45** 6056