Replica-symmetry breaking in perceptrons

1996 J. Phys. A: Math. Gen. 29 3063

(http://iopscience.iop.org/0305-4470/29/12/014)

J. Phys. A: Math. Gen. **29** (1996) 3063–3073. Printed in the UK

Replica-symmetry breaking in perceptrons

W Whyte and D Sherrington†
Theoretical Physics, Department of Physics, 1 Keble Road, Oxford OX1 3NP, UK

Received 5 June 1996

Abstract. In the problem of optimization of pattern stabilization in perceptrons the replica-symmetric ansatz is known to be mathematically unstable for storage capacities α greater than some α_c. In this paper we demonstrate that for α greater than α_c the one-step replica-symmetry broken (RSB) solution is also unstable. We further show that in this region, full RSB is necessary for an exact solution. Direct evaluation of the two-step RSB solution yields a minimum storage error which is only slightly greater than the one-step RSB, which itself is greater than that given by the (unstable) replica-symmetric ansatz by a much larger amount.

1. Introduction

A central study in the field of statistical mechanics of neural networks is the use of the perceptron as an associative memory. A perceptron is a device which associates an output S_o with an input vector $\{S_i\}$, $i = 1 \ldots N$, by a rule of the form $S_o = f(\boldsymbol{J} \cdot \boldsymbol{S})$. In the case of the binary perceptron, $f(\boldsymbol{J} \cdot \boldsymbol{S}) = \mathrm{sign}(\boldsymbol{J} \cdot \boldsymbol{S})$, and the problem of interest is to store a set of p patterns $\boldsymbol{\xi}^\mu$ by finding a synaptic vector $\boldsymbol{J}$ such that for as many of the ξ^μ's as possible the statement

$$\xi^\mu_{\text{output}} = \mathrm{sign}(\boldsymbol{J} \cdot \boldsymbol{\xi}^\mu) \tag{1}$$

is true. It is hoped that the perceptron which stores the patterns correctly will then give the correct output not just when the input corresponds to a stored pattern but also when the input is a corrupted or noisy version of one of the stored patterns; this is known as associativity. In order to improve the associativity one introduces a more stringent condition on the stored patterns; for a given $\kappa > 0$, we now attempt to find a $\boldsymbol{J}$ that satisfies

$$\xi^\mu_{\text{output}} = \mathrm{sign}(\boldsymbol{J} \cdot \boldsymbol{\xi}^\mu - \kappa) \tag{2}$$

for as many patterns as possible (to avoid satisfying this criterion by a simple scaling of $\boldsymbol{J}$, we introduce the condition $\boldsymbol{J} \cdot \boldsymbol{J} = N$). Although fewer patterns may now be stored so as to satisfy this condition, it is expected that those patterns now stored will be more robust against corruption of the input data, giving the desired improvement in associativity.

The behaviour of the network depends on the storage ratio $\alpha \equiv p/N$ and the parameter κ. For any value of κ, there exists an $\alpha_c(\kappa)$ such that the network can store all the patterns correctly only if $\alpha < \alpha_c$. For $\alpha > \alpha_c$ the interesting question is to determine what fraction of the patterns can be stored correctly.

This calculation is typically carried out using the well known 'replica method'. To obtain results using this method, one must make an ansatz for the form of the replica solution as

† Corresponding author.

0305-4470/96/123063+11$19.50

the number of replicas tends to 0. Gardner and Derrida [1] considered the replica-symmetric (RS) ansatz and showed it to be stable against RSB fluctuations for $\alpha < \alpha_c$. However, for $\alpha > \alpha_c$ the RS solution is unstable [1, 2] and RSB is necessary. This result has been borne out by numerical studies by Majer *et al* [3] and Erichsen *et al* [4]. Both of these numerical studies commented on the necessity for a transition to further levels of replica-symmetry breaking, as shown by the incorrect behaviour of various parameters as $\alpha \to \infty$, but neither investigated at what value of α this transition occurred.

In this paper, we investigate the stability of the one-step replica-symmetry broken solution (one-step RSB or RSB1) and show that the argument which demonstrates the instability of this solution can be generalized to demonstrate the instability of any finite level of replica-symmetry breaking. We provide numerical evidence for this by showing that when $\alpha > \alpha_c$ and the number of replicas goes to zero, the two-step RSB ansatz satisfies the replica saddle-point equations better than the RSB1 ansatz. In general, however, the difference in the free energies is $\mathcal{O}(10^{-2})$ compared to the difference between the free energies for replica symmetry and one-step RSB, implying that for most practical purposes one-step RSB will be adequate.

2. The model and the one-step RSB solution

The system we are interested in is a perceptron with N input nodes (labelled $i = 1, \ldots, N$) and one output node, obeying the update rule

$$S_{\mathrm{output}} = \mathrm{sign}\left(\sum_i J_i S_i\right) \qquad S_i \in \{\pm 1\} \tag{3}$$

with the $\{J_i\}$ constrained by the spherical rule $\sum_i J_i^2 = N$. This perceptron is trained to store correctly as many as possible of an ensemble of $p = \alpha N$ patterns $\{\boldsymbol{\xi}^\mu\}$, drawn at random from $\{\pm 1\}^N$. We define the 'aligning field' λ^μ of a pattern $\boldsymbol{\xi}^\mu$ by

$$\lambda^\mu = \xi^\mu_{\mathrm{output}} \sum_i J_i \xi_i^\mu / |J| \,. \tag{4}$$

For some κ, usually taken to be $\geqslant 0$, a pattern $\boldsymbol{\xi}^\mu$ is considered to be stored correctly if its aligning field $\lambda^\mu > \kappa$.

We formulate the problem as 'energy' minimization, defining the energy as

$$E = \sum_\mu g(\lambda^\mu) \tag{5}$$

and, wherever possible, present formal results in terms of a general cost function $g(\lambda)$. For the present explicit case, however, the cost function is

$$g(\lambda) = \tilde{g}(\lambda) \equiv \theta(\kappa - \lambda) \,. \tag{6}$$

The minimized energy therefore gives the minimum possible number of patterns that can be stored incorrectly, in the sense defined above. This is zero if α is less than a critical storage capacity $\alpha_c(\kappa)$.

We calculate this quantity by obtaining the free energy of the system,

$$\begin{aligned} f(\{\xi\}) &= -\lim_{N\to\infty} \frac{1}{N\beta} \ln Z \\ &= -\lim_{N\to\infty} \frac{1}{N\beta} \ln \int \prod_{i=1}^N \mathrm{d}J_i \, \delta((\boldsymbol{J})^2 - N) \, \mathrm{e}^{-\beta E} \end{aligned} \tag{7}$$

and finally taking the limit $\beta \to \infty$ (in which $E = Nf$). The free energy in this limit gives the fraction of patterns that are stored incorrectly, and we thus identify it with the output error of the perceptron. In the usual fashion, the free energy is assumed to be self-averaging over the disorder in the patterns $\{\xi^\mu\}$. In order to perform this average, we employ the replica trick $\langle \ln Z\rangle = \lim_{n\to 0}(\langle Z^n\rangle - 1)/n$.

We now review some known results for this problem [1–4]. Using standard techniques [1, 5] one can express $\langle Z^n\rangle$ in the form

$$\begin{aligned}\langle Z^n\rangle &= \int \prod_{\alpha\beta} \mathrm{d}q_{\alpha\beta}\, \mathrm{e}^{N\Phi(q_{\alpha\beta})} \\ &= \int \prod_{\alpha\beta} \mathrm{d}q_{\alpha\beta} \prod_{\alpha\beta} \mathrm{d}F_{\alpha\beta} \prod_{\alpha} \mathrm{d}E_\alpha \\ &\qquad \times \exp\left[N\left(\alpha G_\Lambda(\{q_{\alpha\beta}\}) + G_J(\{E_\alpha\},\{F_{\alpha\beta}\}) - \sum_{\alpha<\beta} F_{\alpha\beta}q_{\alpha\beta}\right)\right]\end{aligned} \tag{8}$$

where

$$\begin{aligned}\exp[\alpha G_\Lambda(\{q_{\alpha\beta}\})] &\equiv \int \prod_\alpha \mathrm{d}\lambda_\alpha \prod_\alpha \mathrm{d}x_\alpha \exp\left[\sum_\alpha \left(\beta g(\lambda_\alpha) + \mathrm{i}x_\alpha\lambda_\alpha - \tfrac{1}{2}x_\alpha^2\right) - \sum_{\alpha\beta} q_{\alpha\beta}x_\alpha x_\beta\right] \\ \times \exp[G_J(\{E_\alpha\},\{F_{\alpha\beta}\})] &\equiv \int \prod_\alpha J_\alpha \exp\left[\sum_\alpha E_\alpha(1-J_\alpha^2) + \sum_{\alpha\beta} F_{\alpha\beta}J_\alpha J_\beta\right].\end{aligned} \tag{9}$$

Here $q_{\alpha\beta} \equiv \frac{1}{N}\sum_i J_i^\alpha J_i^\beta$ is a measure of the similarity of two replicated networks that both minimize the free energy, and $F_{\alpha\beta}$ is a conjugate variable to $q_{\alpha\beta}$.

We solve equation (8) by the saddle-point method in the limit $N \to \infty$. In order to do this we need to make an ansatz about the form of the solution. The simplest ansatz is replica symmetry (RS), in which we take $E_\alpha = E\ \forall\alpha$ and $q_{\alpha\beta} = q,\ F_{\alpha\beta} = F\ \forall\alpha \neq \beta$. This has been shown [2–4] to be the correct solution to the saddle-point equations for $\alpha \leqslant \alpha_c$; however, for the perceptron above saturation the replica-symmetric solution is unstable and a different ansatz must be used. The simplest non-replica-symmetric ansatz for the solution of equations (7), (8) is given by one-step replica-symmetry breaking (one-step RSB or RSB1) [6] in which we assume the matrix $q_{\alpha\beta}$ has a block structure, with blocks of size $m \times m$ such that the diagonal blocks have 1 in their diagonal entries and q_1 on their off-diagonal, while the off-diagonal blocks have q_0 in all their entries. In other words

$$\begin{aligned} q_{\alpha\beta} &= 1 && \alpha = \beta \\ q_{\alpha\beta} &= q_0 && \text{if the integer parts of } \alpha/m \text{ and } \beta/m \text{ are the same} \\ q_{\alpha\beta} &= q_1 && \text{otherwise}. \end{aligned} \tag{10}$$

Majer *et al* [3] have shown that within the RSB1 subspace, the pattern-choice averaged minimum energy for any cost function $g(\lambda)$ in the limit $\beta \to \infty, q_0 \to 1$ is

$$\begin{aligned} e = \frac{\langle E_{\min}\rangle}{N} = \lim_{\beta\to\infty}\langle f\rangle = \lim_{\beta\to\infty}\max_{\gamma,q_1,w}\Bigg(&\frac{q_1}{2\gamma(1+w\Delta q)} + \frac{\ln(1+w\Delta q)}{2w\gamma} \\ &+ \frac{\alpha}{w\gamma}\int \mathrm{D}z_1 \ln\int \mathrm{D}z_0\, \exp\left(-w\gamma\left[g(\lambda_0) + \frac{(\lambda_0 - z_1\sqrt{q_1} - z_0\sqrt{\Delta q})^2}{2\gamma}\right]\right)\Bigg)\end{aligned} \tag{11}$$

where $\gamma \equiv \beta(1-q_0)$, $w \equiv \beta m/\gamma$ and λ_0 minimizes the expression in the square bracket for given values of $z_0, z_1, \gamma, q_1, \Delta q \equiv (1-q_1)$. If we write $z = z_1\sqrt{q_1} + z_0\sqrt{\Delta q}$, this minimization requires us to invert the function $z(\lambda) = \gamma g'(\lambda) + \lambda$. For the perceptron above saturation, this inverse function is multiple-valued and we are required to perform a Maxwell's construction in order to make it single-valued. This, in turn, gives a discontinuity in $\lambda_0(z)$, whose effects on the stability of the RSB1 solution will be discussed later.

Although the problem was formulated in terms of energy minimization, the solution (11) instead requires us to maximize a quantity. This is due to the fact that in the limit $n \to 0$ the number $n(n-1)$ of independent elements of the matrix $q_{\alpha\beta}$ becomes negative. A given level of RSB will therefore be unstable if the next level of RSB gives a higher value for the minimum energy, and therefore the minimum number of errors.

The maximization with respect to γ, q_1, w in (11) must be performed numerically. This has been done by [3, 4], who have found that whenever the perceptron is above saturation the RSB1 solution gives a higher value for the minimum error than the RS solution, as might be expected from the fact [2] that the RS solution is unstable throughout this region. We now take the work of these studies a step further by investigating the stability of the RSB1 solution, analytically and numerically; the analysis suggests strongly that full RSB is necessary whenever the perceptron is above saturation.

3. Stability of the one-step RSB solution

We now look at the fluctuations of the RSB1 solution around the saddle-point values. A change in the stability of the RSB1 solution is indicated by a change in the sign of one of the eigenvalues of the matrix of the quadratic fluctuations at the saddle point [1, 7]. We define this matrix $\boldsymbol{H}$ as follows: the submatrices $H^{qq}_{(\alpha\beta)(\gamma\delta)}$ and $H^{Eq}_{\alpha(\beta\gamma)}$ are defined by

$$H^{qq}_{(\alpha\beta)(\gamma\delta)} \equiv \frac{\partial^2\Phi}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}} \qquad H^{Eq}_{\alpha(\beta\gamma)} \equiv \frac{\partial^2\Phi}{\partial E_\alpha \partial q_{\beta\gamma}}\,. \tag{12}$$

The other submatrices, $H^{FF}_{(\alpha\beta)(\gamma\delta)}$, $H^{qF}_{(\alpha\beta)(\gamma\delta)}$, $H^{FE}_{(\alpha\beta)\gamma}$, $H^{EE}_{\alpha\beta}$ etc, are defined analogously. The evaluation of the eigenvalues is complicated; however, we can make use of several simplifying arguments, used originally by Dorotheyev [8] for investigating the stability of the RSB1 solution for a pseudo-inverse synaptic matrix. First, we observe that the submatrices are invariant under the action of the hierarchical tree (HT) group. This group is defined in appendix 1 of [8] and in [9]. We may represent it as

$$S_{n/m}\hat{\otimes}(S_m)^{\otimes n/m} \tag{13}$$

with $(S_m)^{\otimes n/m}$ being the direct product of the k-element permutation group with itself n/m times, and $\hat{\otimes}$ being the semidirect product. We can thus express the eigenvectors in terms of the bases of the irreducible representations of the HT group. We further reduce the number of calculations necessary by using the fact that instability only arises in the direction of the replicon-like eigenvectors. In RSB1 there are four families of these eigenvectors, which we call $R^{(a)}$, $R^{(e)}_1$, $R^{(e)}_2$, $R^{(e)}_3$; their definition and the calculation of the corresponding eigenvalues $\gamma_q^{R^{(a)}}, \gamma_F^{R^{(a)}}$, etc are given in the appendix. In order for the RSB1 solution to be stable, the sign of each eigenvalue of the matrix

$$\begin{pmatrix} \alpha\gamma_q^R & 1 \\ 1 & \gamma_F^R \end{pmatrix} \tag{14}$$

must be the same as its sign for a known stable solution, for any $R \in \{R^{(a)}, R^{(e)}_1, R^{(e)}_2, R^{(e)}_3\}$. We know that the RSB1 solution is stable in the limit $\alpha \to 0$, in which case all the eigenvalues in question are -1. We therefore need to find at what point any eigenvalue changes sign, or in other words where

$$\alpha\gamma_q^R\gamma_F^R \geqslant 1\,. \tag{15}$$

We now show that this occurs at the critical storage capacity. Our proof for RSB1 parallels that of Bouten [2] for the instability of the RS solution.

We consider the family $R^{(a)}$, whose eigenvalues, following the notation used in the appendix, can be written as

$$\begin{aligned}\gamma^{R^{(a)}} &= K_1 - 2K_2 + K_3\\ &= \Big[\Big[\big(\big[f^2\big]_f - [f]_f^2\big)^2\Big]_0\Big]_1\,.\end{aligned} \tag{16}$$

Noting that

$$\frac{\mathrm{d}}{\mathrm{d}z_1}\big([1]_f^{m-1}[f]_f\big) = -\mathrm{i}\sqrt{f_1}\big((m-1)[1]_f^{m-2}[f]_f^2 + [1]_f^{m-1}\big[f^2\big]_f\big) \tag{17}$$

where $f_1 = q_1$ if $f = x$ and $f_1 = F_1$ if $f = J$, it is easy to see that in the limit $m \to 0$,

$$\begin{aligned}\gamma_y^{R^{(a)}} &= -\frac{1}{f_1}\left[\frac{\int \mathrm{D}z_0\big(\frac{\mathrm{d}}{\mathrm{d}z_1}[1]_f^{m-1}[f]_f\big)^2[1]_f^{-m}}{\int \mathrm{D}z_0[1]_f^m}\right]\\ &= \frac{1}{m^2 f_1^2}\left[\frac{\int \mathrm{D}z_0\big(\frac{\mathrm{d}^2}{\mathrm{d}z_1^2}[1]_f^m\big)^2[1]_f^{-m}}{\int \mathrm{D}z_0[1]_f^m}\right]\end{aligned} \tag{18}$$

where $f = x$ for $y = q$ and $f = J$ for $y = F$. Using these, it is straightforward to obtain $\gamma_F^{R^{(a)}} = (1-q_0)^2$ and hence

$$\alpha\gamma_F^{R^{(a)}}\gamma_q^{R^{(a)}} = \frac{\alpha}{w^2 q_1^2}\left[\frac{\int \mathrm{D}z_0\big(\frac{\mathrm{d}^2}{\mathrm{d}z_1^2}[1]_x^m\big)^2[1]_x^{-m}}{\int \mathrm{D}z_0[1]_x^m}\right]. \tag{19}$$

The proof of instability above the critical storage capacity follows from there being a discontinuity in $\mathrm{d}/\mathrm{d}z_1([1]_x)$. From the appendix,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z_1}[1]_x &= \frac{1}{2}\frac{\mathrm{e}^G}{(1-\gamma g''(\lambda_0))^{3/2}}\gamma g'''(\lambda_0)\frac{\mathrm{d}\lambda_0}{\mathrm{d}z_1}\\ &\quad + \frac{\mathrm{e}^G}{(1-\gamma g''(\lambda_0))^{3/2}}\left(-w\gamma(g'(\lambda_0)\frac{\mathrm{d}\lambda_0}{\mathrm{d}z_1} - \frac{1}{\gamma}\frac{\mathrm{d}}{\mathrm{d}z_1}(\lambda_0 - \sqrt{q_1}z_1 + \sqrt{\Delta q}z_0)\right)\end{aligned} \tag{20}$$

where $G \equiv w(\gamma g(\lambda_0) - (\lambda_0 - \sqrt{q_1}z_1 + \sqrt{\Delta q}z_0)^2/2)$. For the cost function $\tilde{g}(\lambda)$ (equation (6)) under consideration here, λ_0 is easy to evaluate as a function of $t \equiv \sqrt{q_1}z_1 + \sqrt{\Delta q}z_0$:

$$\begin{aligned}\lambda_0 &= t &&\text{for}\quad t < \kappa - \sqrt{2\gamma}\\ \lambda_0 &= \kappa &&\text{for}\quad \kappa - \sqrt{2\gamma} < t < \kappa\\ \lambda_0 &= t &&\text{for}\quad \kappa < t\,.\end{aligned} \tag{21}$$

Since for $\alpha > \alpha_\mathrm{c}$ there is a discontinuity in λ_0 at $t = \kappa - \sqrt{2\gamma}$, its first and therefore its second derivative contain a delta function at this point. Since this delta function is then squared, it will contribute an infinite positive weight to $\gamma_q^{R_2^{(a)}}$. Therefore, whenever there is a discontinuity in λ_0, which is the case in the entire region above saturation, the sign of $\alpha\gamma_F^{R^{(a)}}\gamma_q^{R^{(a)}}$ is positive. We can therefore say that the the RSB1 ansatz is unstable.

Let us now consider qualitatively the situation for further levels of replica-symmetry breaking. These will still produce a Hessian matrix that is invariant under some generalization of the HT group; the rth level of RSB will have some replicon-like eigenvalue that can be expressed as

$$\gamma_q = c\left[\cdots\left[\left[\frac{\int \mathrm{D}z_0\left(\frac{\mathrm{d}^2}{\mathrm{d}z_1^2}[1]_x^m\right)^2[1]_x^{-m}}{\int \mathrm{D}z_0[1]_x^m}\right]\right]_1\cdots\right]_r \tag{22}$$

to which a generalization of the previous argument can be applied, showing that any replica-symmetry broken solution in which this term has a finite weight will be unstable. This will be the case if there is only a finite degree of RSB. We therefore conclude that for a perceptron above saturation the only exact solution is given by full replica-symmetry breaking. A numerical study of the full replica-symmetry broken solution is, of course, outside the scope of this paper; however, in the next section we perform a numerical study of the two-step RSB solution, demonstrating that throughout the region above saturation it gives a higher minimum number of errors than the RSB1 solution and providing confirmation of the RSB1 result of this section.

4. Two-step replica-symmetry breaking

Two-step replica-symmetry breaking (RSB2) is a relatively straightforward, but numerically complicated, extension of RSB1 [6]. We change our notation in line with the convention introduced in the appendix, so

$$\begin{aligned}&q_{\alpha\beta} = q_{[\alpha_1,\alpha_2,\alpha_3],[\beta_1,\beta_2,\beta_3]}\\ &\alpha_1,\beta_1 = 1,\ldots,n/m_1 \qquad \alpha_2,\beta_2 = 1,\ldots,m_1/m_2 \qquad \alpha_3,\beta_3 = 1,\ldots,m_2\,.\end{aligned} \tag{23}$$

Under this notation, the RSB2 ansatz becomes

$$\begin{aligned}
q_{[\alpha_1,\alpha_2,\alpha_3],[\beta_1,\beta_2,\beta_3]} &= 1 && \text{if } \alpha_1=\beta_1 \text{ and } \alpha_2=\beta_2 \text{ and } \alpha_3=\beta_3\\
q_{[\alpha_1,\alpha_2,\alpha_3],[\beta_1,\beta_2,\beta_3]} &= q_0 && \text{if } \alpha_1=\beta_1 \text{ and } \alpha_2=\beta_2 \text{ and } \alpha_3\neq\beta_3\\
q_{[\alpha_1,\alpha_2,\alpha_3],[\beta_1,\beta_2,\beta_3]} &= q_1 && \text{if } \alpha_1=\beta_1 \text{ and } \alpha_2\neq\beta_2\\
q_{[\alpha_1,\alpha_2,\alpha_3],[\beta_1,\beta_2,\beta_3]} &= q_2 && \text{if } \alpha_1\neq\beta_1\,.
\end{aligned}$$

Considering the general cost function (5) and using standard techniques [1, 3, 5] we can derive the following expression for the averaged minimum error in the limit $\beta\to\infty$, $q_0\to 1$:

$$\begin{aligned}
e = \lim_{\beta\to\infty}\max_{\gamma,q_1,q_2,w_1,w_2}\Bigg[&\frac{q_2}{2\gamma(1+w_1\Delta_0+w_2\Delta_1)}\\
&+\frac{\ln(1+w_1\Delta_0)}{2w_1\gamma}+\frac{\ln[1+w_2\Delta_1/(1+w1\Delta_0)]}{2w_2\gamma}\\
&+\frac{\alpha}{w_2\gamma}\int\mathrm{D}z_2\ln\int\mathrm{D}z_1\Bigg[\int\mathrm{D}z_0\\
&\times\exp\left(-w_1\gamma\left[g(\lambda_0)+\frac{(\lambda_0-z_2\sqrt{q_2}-z_1\sqrt{\Delta_1}-z_0\sqrt{\Delta_0})^2}{2\gamma}\right]\right)\Bigg]^{w_2/w1}\Bigg].
\end{aligned} \tag{24}$$

Here $\gamma = \beta(1 - q_0)$ as before and $w_i \equiv \beta m_i/\gamma$; we also use the shorthand $\Delta_0 = q_0 - q_1, \Delta_1 = q_1 - q_2$. It can easily be seen that if we take $q_1 = q_2$, $\Delta_1 = 0$ and this formula reduces to the RSB1 case.

In the case under investigation here, with the cost function given by (6), the innermost integral reduces to

$$I_{z_0} = \frac{\exp\left(\frac{1}{2}\frac{-w_1A^2}{1+w_1\Delta_0}\right)}{\sqrt{(1+w_1\Delta_0)}}\left[H\left(\frac{A-\sqrt{2\gamma}(1+w\Delta_0)}{\sqrt{\Delta_0(1+w_1\Delta_0)}}\right) - H\left(\frac{A}{\sqrt{\Delta_0(1+w_1\Delta_0)}}\right)\right]$$
$$+\exp(-w_1\gamma)H\left(\frac{\sqrt{2\gamma}-A}{\sqrt{\Delta_0}}\right) + H\left(\frac{A}{\sqrt{\Delta_0}}\right) \tag{25}$$

where $A = \kappa - z_2\sqrt{q_2} - z_1\sqrt{\Delta_1}$ and $H(x) = \int_x^\infty \mathrm{D}u$. We evaluated (24) for a range of values of α and for $\kappa = 1$, performing the maximization numerically. The results are shown in figures 1–5. The transition from RSB1 to RSB2 causes the output error e to increase by an amount that is typically $\mathcal{O}(10^{-4})$; this is small, but greater than the numerical tolerance of our minimization procedure which is $\mathcal{O}(10^{-6})$. Any lack of smoothness in the curves presented is due to a combination of two factors: first, the 'valleys' for this problem prove to be relatively wide; second, the function varies much faster in general with respect to γ, q_1, q_2 than w_1, w_2, meaning that the choice of starting values of w_1, w_2 is of great importance.

We find that throughout the region of replica-symmetry instability the RSB2 solution gives a greater value for the minimum error than the RSB1 solution, showing that RSB1 is not the true solution anywhere in this region. Figure 1 shows the minimum errors evaluated within RSB1 and RSB2 and figure 2 shows their differences. The other figures show the behaviour of the maximizing values of γ, q_1, q_2, w_1, w_2, with the corresponding RSB1 values for comparison. As $\gamma \equiv \lim_{\beta\to\infty}\beta(1-q_0)$, and as we expect that q_0 will increase with increasing levels of RSB, γ^{rsb2} should be less than γ^{rsb1}; our results (figure 3) confirm

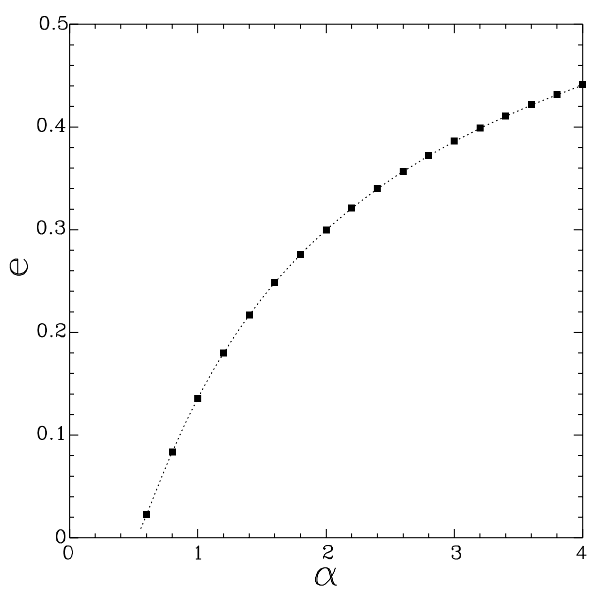

Figure 1. The minimum error e for RSB1 (dotted curve) and RSB2 (full squares), for $\kappa = 1$.

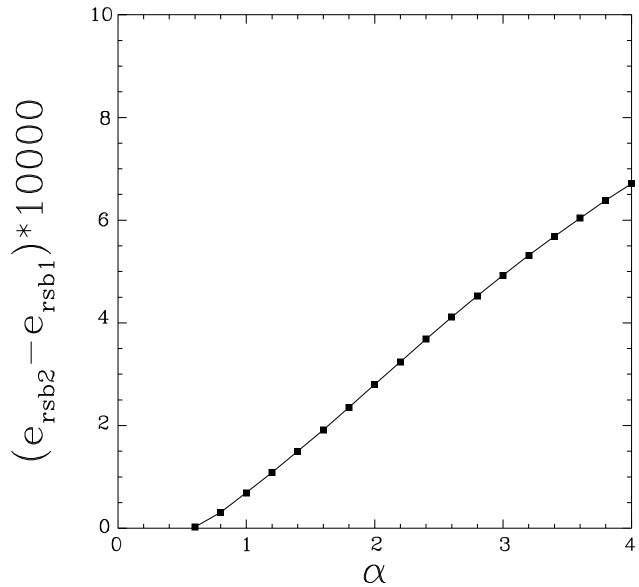

Figure 2. The difference between the minimum errors obtained for the RSB1 and the RSB2 solution, measured in units of 10^{-4}. The curve is a guide to the eye. $\kappa = 1$.

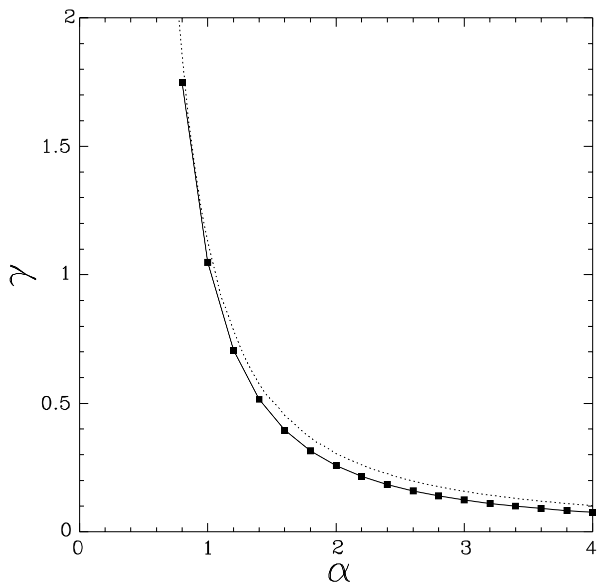

Figure 3. γ^{rsb1} (dotted curve) and γ^{rsb2} (full squares). The full curve is a guide to the eye. $\kappa = 1$.

this. A decrease in γ corresponds to a slight narrowing of the gap in the distribution of local stabilities, which is of magnitude $\sqrt{2\gamma}$ [1, 3]. We have calculated this distribution but the result is sufficiently similar to the RSB1 result that we do not reproduce it here. We find that q_1 and q_2 are respectively higher and lower than q_1^{rsb1} (figure 4), and that $w_1\gamma$ and $w_2\gamma$ are, respectively, higher and lower than $w_1^{\mathrm{rsb1}}\gamma^{\mathrm{rsb1}}$ (figure 5), as would be expected.

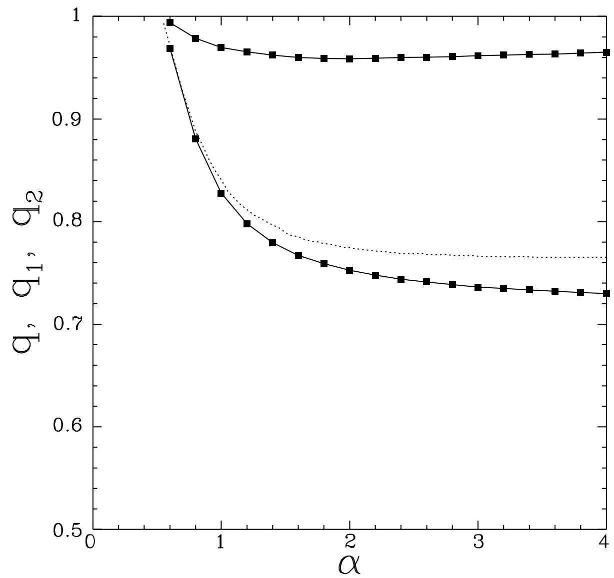

Figure 4. q_1^{rsb1} (dotted curve), q_1^{rsb2} (upper set of points) and q_2^{rsb2} (lower set of points). The full curves are guides to the eye. $\kappa = 1$.

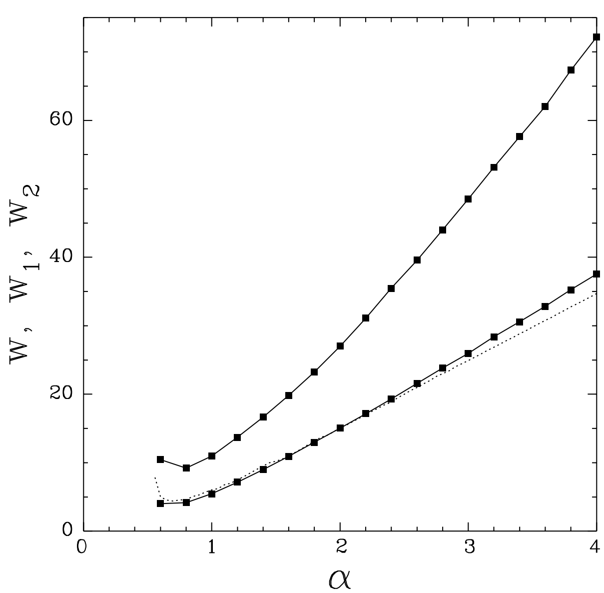

Figure 5. w_1^{rsb1} (dotted curve), w_1^{rsb2} (upper set of points) and w_2^{rsb2} (lower set of points). The full curves are guides to the eye. $\kappa = 1$.

5. Conclusions

It has been known for some time that some degree of replica-symmetry breaking is necessary for an exact solution to the problem of minimum-error storage of patterns in a perceptron above saturation. We have shown that RSB1 is inadequate and suggest that an exact solution requires full RSB. Our results imply that any finite level of replica-symmetry breaking for any cost function $g(\lambda)$ will be unstable if the function $\lambda_0(z, q_i, \gamma)$ contains a discontinuity. The result for the perceptron is backed up by a numerical investigation of the global stability of the RSB1 solution, showing that it will not maximize the minimum error anywhere in the region above saturation. This numerical study has, however, also shown that the observable effects of the transition from RSB1 to RSB2 are very small compared to the effects of the transition from RS to RSB1, so that for most purposes RSB1 may be considered sufficient for this case.

This paper has concentrated on the free energy of the perceptron trained with the Gardner–Derrida cost function. This cost function is somewhat special, in that the discontinuity in λ_0 is present throughout the region $\alpha > \alpha_c$. For cost functions such as the perceptron cost function, $g(\kappa - \lambda) = (\kappa - \lambda)\theta(\kappa - \lambda)$, this is not the case, and the transition to RSB2 is still an open question.

Acknowledgments

WW would like to thank the Maths Department, Trinity College, Dublin, for use of its computer facilities while revising this paper, and the British Council for financial support.

Appendix

This discussion in large part reproduces and expands on the results of [8], making those changes necessary to make it relevant to the present case. By consideration of the group orbits, we can obtain the different matrix elements of the submatrices H^{qq}, H^{FF}. We change our notation as follows:

$$q_{\alpha\beta} = q_{[\alpha_1,\alpha_2],[\beta_1,\beta_2]} \tag{A1}$$

where α_1, β_1 label the blocks and α_2, β_2 label the individual matrix elements within the blocks. Under this notation, the RSB1 ansatz becomes

$$\begin{aligned} q_{[\alpha_1,\alpha_2],[\beta_1,\beta_2]} &= 1 && \text{if } \alpha_1 = \beta_1 \text{ and } \alpha_2 = \beta_2 \\ q_{[\alpha_1,\alpha_2],[\beta_1,\beta_2]} &= q_0 && \text{if } \alpha_1 = \beta_1 \text{ and } \alpha_2 \neq \beta_2 \\ q_{[\alpha_1,\alpha_2],[\beta_1,\beta_2]} &= q_1 && \text{if } \alpha_1 \neq \beta_1\,. \end{aligned} \tag{A2}$$

We make the corresponding changes in notation for $F_{\alpha\beta}$, $H_{(\alpha\beta)(\gamma\delta)}$. The values of those submatrix elements which are relevant to the calculation of the replicon-like eigenvalues can be expressed as follows (subscripts refer to variables being averaged over, superscripts are exponents):

$$\begin{aligned} H_{[1,1][1,2][1,1][1,2]} &\equiv K_1 \\ &= \left[\left[\left[f^2\right]_f^2\right]_0\right]_1 - \left[\left[[f]_f^2\right]_0\right]_1^2 \\ H_{[1,1][1,2][1,1][1,3]} &\equiv K_2 \\ &= \left[\left[\left[f^2\right]_f [f]_f^2\right]_0\right]_1 - \left[\left[[f]_f^2\right]_0\right]_1^2 \\ H_{[1,1][1,2][1,3][1,4]} &\equiv K_3 \\ &= \left[\left[[f]_f^4\right]_0\right]_1 - \left[\left[[f]_f^2\right]_0\right]_1^2 \\ H_{[1,1][2,1][1,1][2,2]} &\equiv L_1 \\ &= \left[\left[\left[f^2\right]_f\right]_0^2\right]_1 - \left[\left[[f]_f\right]_0^2\right]_1^2 \\ H_{[1,1][2,1][1,1][2,2]} &\equiv L_2 \end{aligned} \tag{A3}$$

$$
\begin{aligned}
&= \left[\left[[f^2]_f\right]_0 \left[[f]_f\right]_0^2\right]_1 - \left[\left[[f]_f\right]_0^2\right]_1^2 \\
H_{[1,1][2,1][1,2][2,2]} &\equiv L_3 \\
&= \left[\left[[f]_f^2\right]_0^2\right]_1 - \left[\left[[f]_f\right]_0^2\right]_1^2 \\
H_{[1,1][2,1][1,1][3,1]} &\equiv L_4 \\
&= \left[\left[[f]_f^2\right]_0^2\right]_1 - \left[\left[[f]_f\right]_0^2\right]_1^2 \\
H_{[1,1][2,1][1,2][3,1]} &\equiv L_5 \\
&= \left[\left[[f]_f^2\right]_0 \left[[f]_f\right]_0^2\right]_1 - \left[\left[[f]_f\right]_0^2\right]_1^2 \\
H_{[1,1][2,1][3,1][4,1]} &\equiv L_6 \\
&= \left[\left[[f]_f\right]_0^4\right]_1 - \left[\left[[f]_f\right]_0^2\right]_1^2
\end{aligned}
$$

where $f \equiv x$ for H^{qq}, $f \equiv J$ for H^{FF}, and the averages $[\cdots]_1, [\cdots]_0, [\cdots]_J, [\cdots]_x$ are defined as follows:

$$
\begin{aligned}
&[h]_1 = \int \mathrm{D}z_1\, h(z_1) \\
&\left[\prod_{i=1}^{s} [h_i]_f\right]_0 = \frac{\int \mathrm{D}z_0\, [1]_f^{m-s} \prod_{i=1}^{s} [h_i]_f}{\int \mathrm{D}z_0\, [1]_f^m} \\
&[h]_J = \int \mathrm{d}J\, \mathrm{e}^{(-\frac{1}{2}J^2(2E-F_0)+\mathrm{i}J(\sqrt{F_1}z_1+\sqrt{F_0-F_1}z_0)+E)} h(J) \\
&[h]_x = \int \mathrm{d}\lambda\, \mathrm{d}x\, \mathrm{e}^{(\beta g(\lambda)+\mathrm{i}x(\lambda-\sqrt{q_1}z_1-\sqrt{q_0-q_1}z_0)-\frac{1}{2}(1-q_0)x^2)} h(x)\,.
\end{aligned}
\tag{A4}
$$

The eigenvalues γ^R of the replicon-like eigenvector families R are as follows:

$$
\begin{aligned}
&R^{(a)} : \gamma^{R^{(a)}} = K_1 - 2K_2 + K_3 \\
&R_1^{(e)} : \gamma^{R_1^{(e)}} = L_1 - 2L_2 + L_3 + 2m(L_2 - L_3 - L_4 + L_5) + m^2(L_3 - 2L_5 + L_6) \\
&R_2^{(e)} : \gamma^{R_2^{(e)}} = L_1 - 2L_2 + L_3 \\
&R_3^{(e)} : \gamma^{R_3^{(e)}} = L_1 - 2L_2 + L_3 + m(L_2 - L_3 - L_4 + L_5)\,.
\end{aligned}
\tag{A5}
$$

We can also evaluate $[1]_J$ and (in the limit $\beta \to \infty$) $[1]_x$. These are

$$
\begin{aligned}
&[1]_J = \frac{1}{\sqrt{2E-F_0}} \exp\left(E - \frac{(\sqrt{F_1}z_1 + \sqrt{F_0-F_1})^2}{2(2E-F_0)}\right) \\
&[1]_x = \frac{1}{\sqrt{1-\gamma g''(\lambda_0)}} \exp\left(-w\gamma\left[g(\lambda_0) + \frac{1}{2\gamma}(\lambda_0 - \sqrt{q_1}z_1 - \sqrt{\Delta q}z_0)^2\right]\right).
\end{aligned}
\tag{A6}
$$

References

[1] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
[2] Bouten M 1994 *J. Phys. A: Math. Gen.* **27** 6021
[3] Majer P, Engel A and Zippelius A 1993 *J. Phys. A: Math. Gen.* **26** 7405
[4] Erichsen R Jr and Theumann W K 1993 *J. Phys. A: Math. Gen.* **26** L61
[5] Wong K Y M and Sherrington D 1990 *J. Phys. A: Math. Gen.* **23** 4659
[6] Parisi G 1979 *Phys. Rev. Lett.* **43** 1754; 1980 *J. Phys. A: Math. Gen.* **13** L115, 1101
[7] de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
[8] Dorotheyev E A 1992 *J. Phys. A: Math. Gen.* **25** 5527
[9] Mezard M, Parisi G and Virasoro M 1987 *Spin-Glass Theory And Beyond* (Singapore: World Scientific) p 37