# Replica symmetry instability in perceptron models



1994 J. Phys. A: Math. Gen. 27 6021

(http://iopscience.iop.org/0305-4470/27/17/033)

J. Phys. A: Math. Gen. 27 (1994) 6021–6023. Printed in the UK

COMMENT

# Replica symmetry instability in perceptron models

M Bouten
Limburgs Universitair Centrum, B-3590 Diepenbeek, Belgium

Received 25 April 1994

**Abstract.** It is demonstrated that the replica symmetric saddle point is unstable when the distribution of aligned fields displays a gap.

By applying methods from statistical mechanics within the space of interactions, Gardner and Derrida [1] have developed a powerful new method for determining optimal storage properties of neural network models. Their seminal paper has generated a large number of applications, yielding many quantities of interest for different types of networks. The original form of the method has later been streamlined by Wong and Sherrington [2] and by Griniasti and Gutfreund [3]. The new formalism provides a straightforward procedure for calculating the extremum value within the space of interactions for a large class of cost functions.

Gardner and Derrida (GD) originally applied their method to the determination of the maximum storage capacity $\alpha_c(K)$ of the optimal network for different values of the stability parameter $K$. They moreover determined, for values of the storage ratio $\alpha$ exceeding $\alpha_c(K)$, the minimal fraction of errors that the network will make. The replica symmetry (RS) ansatz was assumed to hold for the saddle point and its local stability was checked. As expected, the RS saddle point was found to be stable for all values of $\alpha$ below the saturation limit $\alpha_c(K)$ where the space of solutions is connected. It was found, however, that local stability continues to hold for larger values of $\alpha$ where one might expect the solution space to be disconnected. GD also determined the Almeida–Thouless line $\alpha_{AT}(K)$ where the stability eventually breaks down. In a recent paper, Majer, Engel and Zippelius [4], while studying the GD problem with a one-stage replica symmetry breaking (RSB) ansatz, made the surprising discovery that the transition from RS to RSB does not occur at $\alpha_{AT}(K)$ but rather at the saturation limit $\alpha_c(K)$ which lies well below the AT line when $K > 0$. This finding made them conclude that RS is *globally* unstable above $\alpha_c(K)$ even where it is *locally* stable.

In the present comment, I want to point out that the AT line calculated by GD and shown in their figure 3 is incorrect. The same figure also appears in [3] and in [4], an indication that the error has gone unnoticed for some time. I will show that the correct AT line coincides exactly with the critical storage line so that $\alpha_{AT}(K) = \alpha_c(K)$. This outcome reduces the surprising finding of Majer, Engel and Zippelius to the normal case.

Following the notation of GD in appendix A2, the condition for local stability of the RS saddle point reads

$$\alpha \gamma_1 \gamma_2 < 1 \qquad (1)$$

where $\gamma_1$ and $\gamma_2$ are given by the GD equations (A2.7) and (A2.13):

$$\gamma_1 = P - 2Q + R = \int_{-\infty}^{\infty} Dt[\overline{X^2}(t) - \overline{X(t)}^2]^2 \tag{2}$$

$$\gamma_2 = P' - 2Q' + R' = (1-q)^2. \tag{3}$$

The functions $\overline{X}(t)$ and $\overline{X^2}(t)$ are given by the GD equations (A2.9) and (A2.10). All these expressions are exact. The wrong result (A2.11) arises from an incorrect calculation of the limit $q \to 1$.

To explain the error and obtain the correct result, it is useful to rewrite (2) in the more convenient form

$$\gamma_1 = -\frac{1}{q}\int_{-\infty}^{+\infty} Dt \left(\frac{\mathrm{d}\overline{X}(t)}{\mathrm{d}t}\right)^2 \tag{4}$$

and to rewrite $\overline{X}(t)$ as

$$\overline{X}(t) = \frac{\mathrm{i}}{1-q}\frac{\int_{-\infty}^{+\infty} \mathrm{d}\lambda\, \exp(-h\theta[K-\lambda] - \frac{1}{2}\frac{(\lambda-\sqrt{q}t)^2}{1-q})(\lambda - \sqrt{q}t)}{\int_{-\infty}^{\infty} \mathrm{d}\lambda\, \exp(-h\theta[K-\lambda] - \frac{1}{2}\frac{(\lambda-\sqrt{q}t)^2}{1-q}}). \tag{5}$$

The limit $h \to \infty$, $q \to 1$ with $h(1-q) = \sigma$ is now easily taken, yielding

$$\overline{X}(t) = \frac{\mathrm{i}}{1-q}[\lambda_0(t,\sigma) - t] \tag{6}$$

where $\lambda_0(t,\sigma)$ is the value of $\lambda$ that minimizes the expressions $\theta[K-\lambda] + \frac{1}{2\sigma}(\lambda - t)^2$. Putting $\overline{X}(t)$ back in (4) to obtain $\gamma_1$ and using the value (3) for $\gamma_2$ yields for the stability condition (1):

$$\alpha\,\gamma_1\,\gamma_2 = \alpha\int_{-\infty}^{+\infty} Dt\left(\frac{\mathrm{d}}{\mathrm{d}t}[\lambda_0(t,\sigma) - t]\right)^2 < 1. \tag{7}$$

This simple expression has been derived previously by Wong and Sherrington [5].

The function $\lambda_0(t,\sigma)$ is easily calculated for the Gardner–Derrida cost function [3]:

$$\lambda_0(t,\sigma) = \begin{cases} t & \text{for } t < K - \sqrt{2\sigma} \\ K & \text{for } K - \sqrt{2\sigma} < t < K \\ t & \text{for } K < t\,. \end{cases} \tag{8}$$

The function $\lambda_0(t,\sigma) - t$ differs from zero only in the interval $K - \sqrt{2\sigma} < t < K$, where it has the constant slope $-1$. Thus one is inclined to write (7) as

$$\alpha\,\gamma_1\,\gamma_2 = \alpha\int_{K-\sqrt{2\sigma}}^{K} Dt < 1 \qquad \text{(WRONG)}. \tag{9}$$

This is exactly the equation (37) in [3] which is equivalent to the GD equation (32). (Note that GD use $x = \sqrt{2\sigma}$.) However, since $\lambda_0(t,\sigma)$ has a discontinuity at $t = K - \sqrt{2\sigma}$, its derivative contains a delta function which, being squared in the integral (7), adds an infinite contribution to $\gamma_1$. This delta function, which dominates the integral (7), has apparently been missed in GD as well as in [3] and [4].

The transition from RS to RSB at $\alpha_c(K)$ is second order, however with one unusual feature. For $\alpha < \alpha_c(K)$, the value of $\alpha\gamma_1\gamma_2$ at zero temperature is smaller than 1 and increases smoothly with growing $\alpha$. At the saturation limit $\alpha_c(K)$, $\alpha\gamma_1\gamma_2$ is still smaller than 1 for all $K > 0$ and is equal to 1 for $K = 0$. Any further increment in $\alpha$, however small, makes the value of $\alpha\gamma_1\gamma_2$ jump abruptly to infinity.

It is straightforward to extend the above result for a general cost function $U(\lambda)$. The sole change occurs in (5), where $\theta[K-\lambda]$ has to be replaced by $U(\lambda)$. The expressions (6) and (7) remain unaltered but the function $\lambda_0(t, \sigma)$ now minimizes the expression $U(\lambda) + \frac{1}{2}(\lambda - t)^2$. Clearly, if $\lambda_0(t, \sigma)$ is a discontinuous function, as in the GD case, condition (7) cannot be satisfied and the RS saddle point is certain to be unstable. It is then necessary to proceed to an RSB calculation. From [2] and [3], it is well known that a discontinuity in $\lambda_0(t, \sigma)$ shows up as a gap in the distribution of the aligned fields [6]. Hence, we obtain the important conclusion that, whenever the distribution of aligned fields displays a gap, the RS saddle point is certainly unstable.

## Acknowledgments

I am indebted to Andreas Engel and Chris Van den Broeck for their insistence that I should write this comment.

*Note added in proof.* Dr W K Theumann has kindly drawn my attention to his paper with R Erichsen (1993 *J. Phys. A: Math. Gen.* **26** L61) in which they observed the transition to RSB at the saturation storage prior to Majer *et al* [4] but likewise failed to notice the error in the calculation of the AT line.

## References

[1] Gardner E and Derrida B 1988 *Phys. A: Math. Gen.* **21** 271
[2] Wong K Y M and Sherrington D 1990 *J. Phys. A: Math. Gen.* **23** 4659
[3] Griniasti M and Gutfreund H 1991 *J. Phys. A: Math. Gen.* **24** 715
[4] Majer P, Engel A and Zippelius A 1993 *J. Phys. A: Math. Gen.* **26** 7405
[5] Wong K Y M and Sherrington D 1993 *Phys. Rev.* E **49** 4465
[6] Wong K Y M and Sherrington D 1990 *J. Phys. A: Math. Gen.* **23** L175

COMMENT

# Reply to the comment of M Bouten

B Derrida

Service de Physique Thèorique, CE Saclay, F91191 Gif sur Yvette, France

Received 15 July 1994

**Abstract.** I agree with the comment of M Bouten that in our paper (1994 *J. Phys. A: Math. Gen.* **27** 6021) the limit $q \to 1$ was incorrect and that equation (A2.11) is wrong.

I agree with the comment of M Bouten [1] that in our paper [2] the limit $q \to 1$ was incorrect and that equation (A2.11) of [2] is wrong. The reason, as pointed out in [1], is that in the integral (A2.7) of [2] a singularity develops for $K - \sqrt{q}t \simeq x$ in the limit $q \to 1$ and $h \to \infty$. This had been missed in [2].

This invalidates the claim [2] that there exists a range above saturation where the replica symmetric solution is stable.

## References

[1] Bouten M 1994 *J. Phys. A: Math. Gen.* **27** 6021

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