Spin-glass phase in a neutral network with asymmetric couplings

1988 J. Phys. A: Math. Gen. 21 L1181

(http://iopscience.iop.org/0305-4470/21/24/004)

J. Phys. A: Math. Gen. 21 (1988) L1181-L1186. Printed in the UK

LETTER TO THE EDITOR

Spin-glass phase in a neural network with asymmetric couplings

R Kree†, D Widmaier‡ and A Zippelius§

† Institut für Theoretische Physik IV, Universität Düsseldorf, D-4000 Düsseldorf, Federal Republic of Germany
‡ Institut für Festkörperforschung der Kernforschungsanlage Jülich, Postfach 1913, D-5170 Jülich, Federal Republic of Germany
§ Institut für Theoretische Physik, Universität Göttingen, Bunsenstrasse 9, D-3400 Göttingen, Federal Republic of Germany

Received 19 August 1988

Abstract. We study the phase diagram of a neural network model which has learnt with the ADALINE algorithm, starting from *tabula non rasa* conditions. The resulting synaptic efficacies are not symmetric under an exchange of the pre- and post-synaptic neuron. In contrast to several other models which have been discussed in the literature, we find a spin-glass phase in the asymmetrically coupled network. The main difference compared with the other models consists of long-ranged Gaussian correlations in the ensemble of couplings.

The dynamics of spin-glass-like systems with asymmetric couplings between the spins has been discussed recently in a number of publications [1-5]. Models of this type are of special interest in the field of neural networks where the couplings J_{ij} correspond to synaptic efficacies, which describe the influence of the presynaptic neuron j on the postsynaptic neuron i. In physiological networks [6] $J_{ij} \neq J_{ji}$, whereas in the standard Hopfield model [7] (for a review, see [8]) the assumption of special learning rules (e.g. Hebb's rule) leads to $J_{ij} = J_{ji}$. Meanwhile, however, other learning rules have been proposed [9, 10], which give rise to asymmetric couplings.

A characteristic feature of networks with symmetric J_{ij} is the appearance of a spin-glass phase [7] with frozen-in magnetic moments, which are not at all or only weakly correlated with any of the learnt patterns. Up to now all the work [1-5], which attempted to introduce asymmetric couplings, arrived at the conclusion that the spin-glass phase is destroyed by a small amount of asymmetry in the J_{ij}.

In the following we give an example of a neural network with asymmetric couplings (arising from a learning rule proposed in [9]) which does show a spin-glass phase. The important difference compared with the other models which have been considered so far is the presence of long-ranged Gaussian correlations in the ensemble of couplings J_{ij}, as will be discussed below.

The couplings are defined by a learning rule of the '*tabula non rasa*' type [9]. The network consists of N neurons and has to learn p patterns $\{\zeta_i^\nu = \pm 1\}$, $\nu = 1, \ldots, p$ and $i = 1, \ldots, N$. The initial values for the synaptic efficacies, $J_{ij}(t=0) = B_{ij}$, are taken to be random and symmetric, i.e. $B_{ij} = B_{ji}$. Learning proceeds with the ADALINE algorithm [11, 12]†, so that the resulting $N \times N$ matrix $\hat{J}$ is given by

$$\hat{J} = J_0 \hat{P} + \hat{B}(\hat{1} - \hat{P}). \tag{1}$$

Here $\hat{P}$ denotes the projector onto the subspace of patterns $\{\zeta_i^\nu\}$. The distribution of B_{ij} is taken to be Gaussian with zero mean and variance $\langle B_{ij}^2 \rangle = B^2/N$. Though the B_{ij} are symmetric and uncorrelated, these properties are *not* shared by the J_{ij}.

The dynamic evolution of our network is defined by a set of Langevin equations for soft spins s_i (graded neurons):

$$\Gamma_0^{-1} \partial_t s_i = -\sum_{j(\neq i)} J_{ij} s_j - f_u(s_i) + \eta_i(t) + h_i^{\text{ext}}(t). \tag{2}$$

The noise $\eta_i(t)$ is a Gaussian variable with zero mean and variance $\langle \eta_i(t) \eta_j(t') \rangle = (2T/\Gamma_0)\delta_{ij}\delta(t-t')$. The microscopic timescale is set by Γ_0^{-1}, h_i^{ext} denotes an external field and $f_u(s_i)$ restricts fluctuations around $|s_i| = 1$. It is chosen such that for large u the fluctuations in spin length are more and more suppressed and for $u \to \infty$ the Ising limit is reached ($s_i = \pm 1$). We note that the soft-spin formulation is not essential. Equivalent results can be achieved for a single spin-flip Glauber dynamics.

We shall only consider the case of a finite number p of patterns, which are taken as independent random variables with zero magnetisation

$$\frac{1}{N} \sum_{i=1}^{N} \zeta_i^\nu = 0.$$

For the purpose of illustration we consider the case of one learnt pattern ζ_i, such that $P_{ij} = (1/N)\zeta_i \zeta_j$. The generalisation to a finite number of patterns will be discussed at the end. For the one-pattern model the explicit dependence on the pattern ζ_i can be gauged away by the transformation

$$\begin{aligned} S_i &\to \zeta_i S_i \\ J_{ij} &\to \zeta_i J_{ij} \zeta_j \end{aligned} \tag{3}$$

which leaves (2) invariant, if $f_u(s_i)$ is odd in s_i. Note that the $\zeta_i \eta_i$ and $\zeta_i B_{ij} \zeta_j$ have the same distribution as η_i and B_{ij}, respectively. The transformed J_{ij} take on the simple form

$$J_{ij} = \frac{J_0}{N} + B_{ij} - \frac{1}{N} \sum_k B_{ik}. \tag{4}$$

We want to elucidate the effects of asymmetric couplings and generalise the model of (4):

$$J_{ij} = \frac{J_0}{N} + B_{ij} - \frac{1}{2N} \sum_k [B_{ik}(1+\lambda) + B_{jk}(1-\lambda)] \tag{5}$$

such that the degree of asymmetry λ can be varied. Note that $\lambda = 1$ corresponds to the original suggestion of [9] whereas $\lambda = 0$ corresponds to a fully symmetrised model.

Using standard techniques [13] it is possible to reduce the N coupled equations of motion to a non-Markovian single-spin dynamics. This has to be solved self-consistently for the overlap

$$m(t) = \left\langle \frac{1}{N} \sum_{i=1}^{N} \langle s_i(t) \rangle_\eta \right\rangle_\zeta \tag{6a}$$

† The convergence for arbitrary initial conditions have been proved by L Personnaz (private communication).

for the autocorrelation

$$C(t, t') = \left\langle \frac{1}{N} \sum_{i=1}^{N} \langle s_i(t) s_i(t') \rangle_\eta \right\rangle_\zeta \tag{6b}$$

and for the response function

$$G(t, t') = \left\langle \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \langle s_i(t) \rangle_\eta}{\partial h_i^{\text{ext}}(t')} \right\rangle_\zeta . \tag{6c}$$

We are interested here in a stationary state, which is characterised by $m(t) = m$, $C(t, t') = C(t-t')$ and $G(t, t') = G(t-t')$. In this case the equation of motion simplifies to

$$-(\mathrm{i}\omega/\Gamma_0)s(\omega) = B^2 G(\omega)s(\omega) + \{J_0 - B^2[1 - \tfrac{1}{4}(1-\lambda^2)]G(\omega = 0)\} m\delta(\omega) + y(\omega) + f_u(s; \omega) + h^{\text{ext}}(\omega) \tag{7}$$

where $f_u(s; \omega)$ is the Fourier transform of $f_u(s(t))$ and the noise $y(\omega)$ is Gaussian distributed with zero mean $\langle y \rangle_y = 0$ and variance

$$\langle y(\omega) y(\omega') \rangle_y = \delta(\omega + \omega')(2T/\Gamma_0 + B^2\{C(\omega) - [1 - \tfrac{1}{4}(1-\lambda)^2] m^2 \delta(\omega)\}). \tag{8}$$

The self-consistency conditions are: $m = \langle s(t) \rangle_y$, $C(t-t') = \langle s(t) s(t') \rangle_y$ and $G(t-t') = \partial \langle s(t) \rangle_y / \partial h^{\text{ext}}(t')$.

A self-consistent solution of equations (7) and (8) is $m = 0$. The equation of motion is then identical to the Sherrington–Kirkpatrick model with relaxational dynamics—no matter how strong the asymmetry λ. Hence we expect a transition from a paramagnetic to a spin-glass phase for small J_0/B. To discuss the general case we split the correlation function into a time-persistent part $q = \lim_{t\to\infty} C(t)$ and a decaying part $\tilde{C}(t) = C(t) - q$. If $q \neq 0$ and/or $m \neq 0$, the noise acquires a static part $y(\omega) = \tilde{y}(\omega) + z\delta(\omega)$ with variances

$$\Delta_z \equiv \langle z^2 \rangle_z = B^2[q - m^2 + m^2 \tfrac{1}{4}(1-\lambda)^2] \tag{9a}$$

and

$$\langle \tilde{y}(\omega) \tilde{y}(\omega') \rangle = \delta(\omega + \omega')(2T/\Gamma_0 + B^2 \tilde{C}(\omega)). \tag{9b}$$

The equation of motion is then

$$-(\mathrm{i}\omega/\Gamma_0)s(\omega) = B^2 G(\omega) s(\omega) + H(z)\delta(\omega) + \tilde{y}(\omega) + f_u(s; \omega). \tag{10}$$

The spin relaxes in a static field

$$H(z) = z + J_0 m - B^2[1 - \tfrac{1}{4}(1-\lambda^2)]G(0)m$$

which has a systematic part $\sim m$ and a fluctuating component z. The modification of the bare propagator is just like in the Gaussian spin glass and so is the dynamic noise, which is generated by the random couplings B_{ij}. In the high-temperature phase ($H = 0$) equations (9) and (10) are consistent with a fluctuation dissipation theorem, which relates correlation (and response) function: $\beta \tilde{C}(\omega) = (2/\omega) \operatorname{Im} G(\omega)$. This can be shown perturbatively in an expansion in the non-linearity of $f_u(s; \omega)$. A non-zero static field $H \neq 0$ gives rise to a non-zero m and q so that the expansion of the non-linearity has to be done around a non-zero $\langle s \rangle$. This does not invalidate the FDT, so that the stationary probability distribution can be constructed for all values of λ. The expectation values of q and m are given in the Ising limit by

$$m = \int \mathrm{D}z \tanh \beta H(z) = K(q, m) \tag{11a}$$

and

$$q = \int \mathrm{D}z \tanh^2 \beta H(z) = F(q, m) \tag{11b}$$

with $\int_{-\infty}^{+\infty} \mathrm{D}z = \int \exp(-z^2/2\Delta_z)\, \mathrm{d}z/(2\pi\Delta_z)^{1/2}$.

The paramagnetic phase becomes unstable to ferromagnetic ordering if $\partial K(q=0, m=0)/\partial m = 1$, i.e. for $J_0 - (B^2/T_{\mathrm{FM}})[1-\frac{1}{4}(1-\lambda^2)] = T_{\mathrm{FM}}$. The paramagnetic phase is unstable to spin-glass fluctuations if $\partial F(q=0, m=0)/\partial q = 1$, i.e. $T_{\mathrm{SG}} = B$. These lines determine the phase boundaries of the paramagnetic phase. The transition line from spin glass to ferromagnet is given by

$$\frac{\partial K(q, m=0)}{\partial m} = 1 = \{J_0^c - \beta B^2(1-q)[1-\tfrac{1}{4}(1-\lambda^2)]\}\beta(1-q).$$

Analytic results can be obtained near the multicritical point and in the limit of low temperature. We find that at $J_0^c/B = \sqrt{\pi/2} + \sqrt{2/\pi}[1-\frac{1}{4}(1-\lambda^2)]$ the ferromagnet and the spin glass exchange their local stability for $T=0$. The magnetisation vanishes at this point so that there is a second-order spin glass to ferromagnetic transition. The phase diagram is shown in figure 1 for $\lambda = 1$ as a function of J_0/B and T/B.

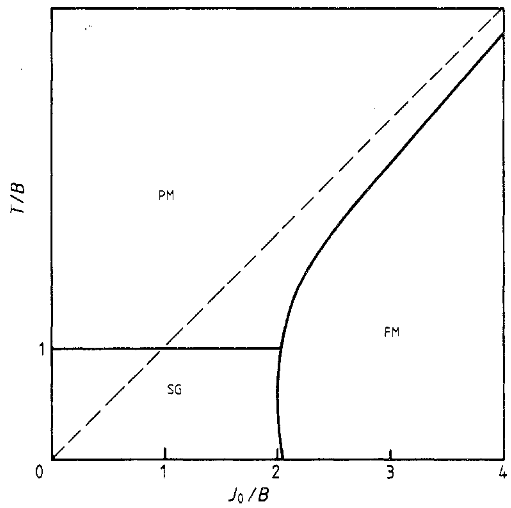

Figure 1. The phase diagram of a neural network model which has learnt with the ADALINE algorithm, starting from *tabula non rasa* conditions, indicating paramagnetic (PM), ferromagnetic (FM) and spin-glass (SG) phases.

Our results are easily generalised to the case of a finite number of patterns. In particular the stationary values of q and m (equation (11)) are unchanged if one restricts oneself to retrieval states.

The stationary probability distribution can be given for arbitrary λ, because the asymmetry only enters into the static field $H(z)$. It is not clear how the relaxation from an arbitrary initial condition proceeds. Starting from a random initial condition, i.e. $m=0$, in the parameter range of the spin-glass phase will result in a stationary state $q \neq 0$. An interesting open problem is the relaxation from an initial state $m(t_0) \neq 0$.

So far we have only discussed the simplest dynamic mean-field theory, which does not show any anomalous response. This solution is dynamically unstable, as expected, in all of the spin-glass phase and the ferromagnetic phase below the de Almeida-Thouless [14] line

$$\left(\frac{T_{\mathrm{dAT}}(\lambda)}{B}\right)^2 = \int \mathrm{D}z \frac{1}{\cosh^4 \beta H(z)}.$$

One of the effects of replica symmetry breaking is an extension of the ferromagnetic phase at the expense of the spin-glass phase. According to an argument of Toulouse [15], we expect the phase boundary between the spin glass and the retrieval phase to be vertical.

We have shown the existence of a spin-glass phase in a neural network with asymmetric couplings. To get some further insight we study the distribution of the coupling constants. The symmetric part $J^s_{ij} = \frac{1}{2}(J_{ij} + J_{ji})$ and the antisymmetric part $J^a_{ij} = \frac{1}{2}(J_{ij} - J_{ji})$ are correlated random variables. Consider, for example, the second cumulant of the antisymmetric part

$$\langle J^a_{ij} J^a_{kl} \rangle_c = \frac{B^2 \lambda^2}{4N^2} (\delta_{ik} + \delta_{jl} - \delta_{il} - \delta_{jk}).$$

The short-range correlations, $\langle (J^a_{ij})^2 \rangle_c = B^2\lambda^2/2N^2$, are indeed negligible as compared to the short-range correlations of the second cumulant of the symmetric part, namely $\langle (J^s_{ij})^2 \rangle_c = \mathrm{O}(1/N)$. Hence our results do not contradict previous studies of spin-glass-like neural networks with short-range correlations only. It is only due to the *long-range* correlations that the dynamics are influenced by J^a_{ij} in the thermodynamic limit. Recently Krauth *et al* [16] conjectured that most of the dynamical effects in neural networks are controlled by two parameters:

$$\eta = \sum_{i \neq j} J_{ij} J_{ji} \left(\sum_{i \neq j} J^2_{ij} \right)^{-1}$$

to characterise the symmetry of the couplings and

$$\Delta_i = \min_{\mu} \zeta^\mu_i J_{ij} \zeta^\mu_j \left(\sum_j J^2_{ij} \right)^{-1}$$

to characterise the stability of the patterns. Our model illustrates that their conjecture should be supplemented by further parameters to specify the long-range correlations of J_{ij}.

One of us (DW) would like to thank M Opper for many helpful discussions. RK acknowledges partial support by the SFB 237 of the Deutsche Forschungsgemeinschaft.

References

[1] Hertz J A, Grinstein G and Solla S A 1986 *Proc. Heidelberg Colloq. on Glassy Dynamics* ed J L van Hemmen and I Morgenstern (Berlin: Springer) p 538
[2] Bausch R, Janssen H K, Kree R and Zippelius A 1986 *J. Phys. C: Solid State Phys.* **19** L779
[3] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167
[4] Crisanti A and Sompolinsky H 1987 *Phys. Rev.* A **36** 1922; 1988 *Phys. Rev.* A **37** 4865
[5] Rieger H, Schreckenberg M and Zittartz J 1988 *J. Phys. A: Math. Gen.* **21** L263

[6] Kandel E R and Schwartz J H 1985 *Principles of Neural Science* (Amsterdam: Elsevier)
[7] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554; 1984 *Proc. Natl Acad. Sci. USA* **81** 3088
[8] Amit D 1986 *Proc. Heidelberg Colloq. on Glassy Dynamics* ed J L van Hemmen and I Morgenstern (Berlin: Springer) p 430
[9] Toulouse G, Dehaene S and Changeux J P 1986 *Proc. Natl Acad. Sci. USA* **83** 1695
[10] Personnaz L, Guyon I, Dreyfus G and Toulouse G 1986 *J. Stat. Phys.* **43** 411
[11] Widrow G and Hoff M E 1960 *Adaptive Switching Circuits, Ire Wescon Conv. Record* part 4, p 96
Diederich S and Opper M 1987 *Phys. Rev. Lett.* **58** 949
[12] Kohonen T 1984 *Self Organization and Associative Memory* (Berlin: Springer)
[13] Bausch R, Janssen H K and Wagner H 1976 *Z. Phys.* B **24** 113
De Dominicis C and Peliti L 1978 *Phys. Rev.* B **18** 353
Sompolinsky H and Zippelius A 1987 *Phys. Rev.* B **25** 6860
[14] de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
[15] Toulouse G 1980 *J. Physique Lett.* **41** 447
[16] Krauth W, Nadal J P and Mézard M 1988 *Preprint*