

Metastable states of neural networks incorporating the physiological Dale hypothesis

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1990 J. Phys. A: Math. Gen. 23 249

(<http://iopscience.iop.org/0305-4470/23/2/020>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 129.252.86.83

The article was downloaded on 01/06/2010 at 09:21

Please note that [terms and conditions apply](#).

Metastable states of neural networks incorporating the physiological Dale hypothesis

T Fukai

Department of Management and Information Science, Gumma Women's College, 501 Nakaohri-cho, Takasaki 370, Japan

Received 25 April 1989

Abstract. Physiologically it is likely that excitatory and inhibitory neurons are rather clearly distinguished or, in other words, each neuron has in most cases a unique excitatory or inhibitory property (the Dale hypothesis). To study the consequence of the physiological constraint, we propose a learning rule for neural networks which incorporates the constraint. Then the distribution of the metastable states is calculated and it is found that the retrieval states form a much larger group in the proposed model than in the Hopfield model. We also study the process of retrieval by considering the statistical dynamics of the overlaps. The result suggests that excitatory neurons and inhibitory neurons are preferably balanced in number if the attraction basins of the stored patterns are to be sufficiently large.

1. Introduction

Apart from fruitful applications of neural networks to engineering or computer scientific problems, attempts have been made to understand the complex performances of the human brain through neural networks (Eccles 1977, Kohonen 1977, Virasoro 1988). If we try to simulate the complex structure or functions of the brain by means of neural networks, the implications of many physiological constraints on biological nervous systems must be understood. There have been various attempts to accommodate physiological constraints into neural network models (Amari 1977a, Clark 1988, Shinomoto 1987, Treves and Amit 1989) though it has not yet been fully clarified whether such physiological constraints play essential roles in biological nervous systems or are simply regarded as neuro-chemically inevitable constraints. Here we study the metastable states of a simple neural network subject to a specific physiological constraint, namely the Dale hypothesis, which claims that each neuron is either excitatory or inhibitory.

A neural network with the Dale hypothesis was studied by Shinomoto (1987) and it was reported that the network acquires a sort of 'cognitive ability'. We examine the effect of incorporating the physiological constraint from different points of view. After proposing a simple neural network model which follows the physiological constraint, we calculate the distribution of the metastable states for the model. The simple realisation of the learning rule to incorporate the constraint makes it possible to carry out the statistical calculations involved. Such distributions were calculated for the Hopfield model (Gardner 1986) and the asymmetrically diluted Hopfield model (Treves and Amit 1988) and it was found that there exist two groups of the metastable states: one forms around the original patterns and the other around spin-glass type configurations. We find that the former group consists of the retrieval states and becomes much

larger in the proposed model than that in the conventional Hopfield (1982) model, which is characteristic of the model incorporating the constraint. Note that the purpose of the present paper is not to construct a realistic model of biological nervous systems but to understand possible effects of the physiological constraint. Therefore we define a minimum model which satisfies the physiological constraint so that we can approach it mostly through analytical calculations.

The paper is organised as follows. In section 2, the learning rule which incorporates the physiological constraint is presented and the model is defined. In section 3, the number of the metastable states is calculated and the results are shown for various cases. Detailed calculations are shown in the appendix. In section 4, the statistical dynamics of overlaps is discussed to see the features of retrieval process. Concluding remarks are given in section 5.

2. The model

We propose a learning rule in which the above physiological constraint is taken into account in a simple manner. We first label each neuron according to its excitatory or inhibitory nature. Let η_j be 1 (-1) if the j th neuron is excitatory (inhibitory). Then the learning of patterns $\{\xi_i^\mu\}$ ($i = 1, \dots, N$, $\mu = 1, \dots, p$) to the synaptic couplings is defined by

$$T_{ij} = \frac{1}{N} \sum_{\mu=1}^p \varepsilon_{ij}^\mu \xi_i^\mu \xi_j^\mu \quad \varepsilon_{ij}^\mu = 1 + \eta_j \xi_i^\mu \xi_j^\mu. \quad (1)$$

Since $\varepsilon_{ij}^\mu \neq 0$ when the product of the signal $\xi_i^\mu \xi_j^\mu$ and η_j is positive, only positive signals are learned by an excitatory neuron. Similarly, only negative signals are learned by an inhibitory neuron. Thus the synaptic couplings generated by (1) satisfy the physiological constraint under consideration.

The learning rule $T_{ij} = 2\theta(T_{ij}^H)T_{ij}^H$, where T_{ij}^H stands for the conventional Hebbian coupling $(1/N)\sum_{\mu} \xi_i^\mu \xi_j^\mu$, was adopted by Shinomoto (1987) to generate the synaptic couplings which respect the physiological constraint. We call this learning rule 'all or nothing learning' in the present paper. Advantages of choosing the present couplings are the following two points. (i) It suits an intuitive image of learning as a sequential process since the synaptic couplings are modified every time when a memory is stored by the network. (ii) T_{ij} is simply written as

$$T_{ij} = T_{ij}^H + \alpha \eta_j$$

with $\alpha = p/N$, which makes any statistical mechanical calculations rather easier. Note that T_{ij} is symmetric if both the i th and the j th neurons are excitatory or inhibitory and otherwise T_{ij} is asymmetric.

The memory retrieval process is an ordinary synchronous process, i.e. spin $S_i(t)$ is updated according to

$$S_i(t+1) = \text{sgn} \left(\sum_{j \neq i} T_{ij} S_j(t) \right). \quad (2)$$

One may think that the above process is not realistic. However, as mentioned previously, we are not aiming at constructing a realistic model. What we are interested in is the characteristic properties of the neural networks whose unit neurons are specified uniquely as excitatory or inhibitory. Therefore we leave the retrieval algorithm unchanged to see only the results of the physiological constraint.

3. Distribution of the metastable states

We calculate the number of the metastable states at the Hamming distance Ng from a pattern $\{\xi_i^r\}$. Since $\{S_i\}$ is a fixed point of the updating algorithm (2) if

$$S_i h_i > 0 \quad h_i = \sum_{j \neq i} T_{ij} S_j \quad (3)$$

is satisfied for any i , the number N_{ms} of the metastable states with the local fields h_i whose sizes are greater than a positive value h is given by

$$N_{ms} = \text{Tr}_s \left\langle \prod_i \theta(S_i h_i - h) \right\rangle_\xi \quad (4)$$

where the trace is taken over the states separated from $\{\xi_i^r\}$ by the Hamming distance Ng and $\langle \dots \rangle$ stands for quenched averages over the random patterns $\{\xi_i^\mu\}$ ($\mu \neq r$). Using the following integral representation of the θ function:

$$\theta(x - a) = \int_{-\infty}^{\infty} du \int_a^{\infty} \frac{d\tau}{2\pi} \exp[iu(\tau - x)]$$

equation (4) can be written as

$$\text{Tr}_s \prod_i \int_{-\infty}^{\infty} dx_i \int_h^{\infty} \frac{d\tau_i}{2\pi} \left\langle \exp \left(i \sum_i x_i \tau_i - i \sum_i x_i S_i h_i \right) \right\rangle_\xi \quad (5)$$

We evaluate the right-hand side of (5) in the limit that N goes to infinity keeping α fixed. We show the details of calculations in the appendix and only present the results obtained. Inserting a parameter γ in front of the second term of ε_{ij}^μ to connect the present model smoothly with the Hopfield model (i.e. $\gamma = 1$ yields the present physiological model while $\gamma = 0$ yields the Hopfield model), we calculate N_{ms} to obtain

$$N_{ms} = \exp(NF(g, \alpha, r, q)) \quad (6)$$

$$\begin{aligned} F(g, \alpha, r, q) = & g(1 - q) \ln(\Phi(s_+)) + (1 - g - r + gq) \ln(\Phi(s_-)) \\ & + gq \ln(\Phi(t_+)) + (r - gq) \ln(\Phi(t_-)) + \alpha [b + \frac{1}{2} \ln a - \frac{1}{2} + (1 - b)^2 / 2a] \\ & - g \ln g - (1 - g) \ln(1 - g) \end{aligned} \quad (7)$$

where

$$\begin{aligned} \Phi(x) &= \int_x^{\infty} \frac{dt}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2) \\ s_{\pm} &= \frac{h \pm (1 - 2g) + \alpha(b - \gamma)}{\sqrt{\alpha a}} \quad t_{\pm} = \frac{h \pm (1 - 2g) + \alpha(b + \gamma)}{\sqrt{\alpha a}} \end{aligned}$$

and saddle points are to be substituted for the two parameters a and b . To derive the above expression, the number of excitatory neurons in the network has been assumed to be rN ($0 < r < 1$) and the number of excitatory neurons taking wrong spin states (i.e. $S_i \neq \xi_i^r$) to be Ngq ($0 < gq < r$). It is necessary to give the distance and the direction of $\{S_i\}$ from $\{\xi_i^r\}$ by g and q , respectively to specify the state $\{S_i\}$.

Equations (6) and (7) yield an upper bound for the number of metastable states (Gardner 1986). We see that the metastable states exist only for the regions of the parameters in which the function F takes positive values. If F takes negative values, there exist no metastable states for such regions of the parameters. Note that the first

four terms in the expression of F give negative contributions while the next two terms give positive contributions. Since $t_{\pm} > s_{\pm}$ for any allowed values of the parameters, we see that F tends to be positive as q becomes smaller.

The function F is evaluated numerically and the results are shown below for various cases. Hereafter h is set equal to 0. We first show the cases of the Hopfield model for comparison. We recovered the results obtained by Gardner (1986). In figure 1(a), F is shown as a function of g when $\alpha = 0.05$ and 0.1. An important point is that there exist two distinct positive bands: one is around $g = 0$ and the other is around $g = \frac{1}{2}$. The former band corresponds to a set of metastable states which are very close to the original pattern but not completely equivalent to it: the closest metastable state in the band has very small but non-vanishing Hamming distance

$$g(\alpha) \approx \left(\frac{\alpha}{2\pi}\right)^{1/2} \exp\left(-\frac{1}{2\alpha}\right)$$

from the original pattern. We may call these states the retrieval states. On the other hand, the later band corresponds to a set of metastable states around the spin-glass type configurations which were found in the thermodynamical calculations (Amit *et al* 1985a, b, 1987) and is much broader than the former.

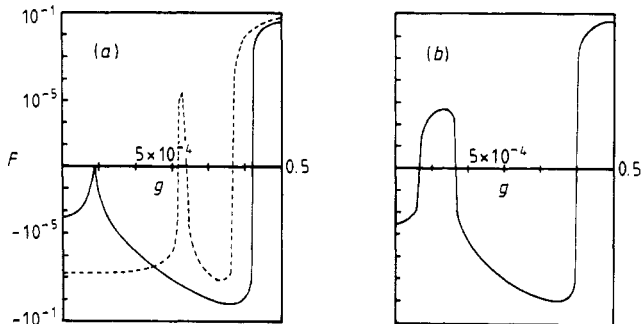


Figure 1. The function F for the Hopfield model (a) and for the present model (b) with equal numbers of excitatory and inhibitory neurons. The number of the metastable states is given by $\exp(NF)$, and therefore the metastable states exist only for the parameter regions in which F takes positive values. In the case of the Hopfield model, the full curve is for $\alpha = 0.05$ and the broken curve is for $\alpha = 0.1$. The height of the former peak is less than $O(10^{-8})$. For the present model, the case when $\alpha = 0.05$ and $q = 0$ is shown.

Now we show the function F for the present model. For the time being, we set q equal to 0 and consider the network with equal numbers of excitatory and inhibitory neurons. In figure 1(b), F is shown when $\alpha = 0.05$. The remarkable point is that the positive band around $g = 0$ becomes much broader and higher than that in the Hopfield model shown in figure 1(a), or equivalently, *the number of retrieval states for the present model increases significantly compared to that for the Hopfield model when $N > O(10^5)$* . This feature is seen for any values of α for which the network behaves as a good memory and is characteristic for the neural network whose unit neurons are uniquely assigned to be excitatory or inhibitory. The immediate implication of this result is that each pattern is stored as a huge set of patterns almost identifiable to the original pattern and consequently identifiable to each other. We will briefly discuss its possible physical or biological implications later. Note also that the number of uncorrelated

metastable states increases slightly. Therefore it is not always true that the number of spurious states decreases for neural networks with asymmetric couplings. (See also Parisi 1986, Feigelman and Ioffe 1986, 1987, Treves and Amit 1988.)

It is known that the gap between the two positive bands disappears if α is larger than some critical value α_c . For the Hopfield model, α_c was estimated to be 0.113 (Gardner 1986). Setting $q = 0$, i.e. observing the direction in which most of the retrieval states exist, we obtain the following value of α_c for the present physiological model when $r = 0.5$:

$$\alpha_c = 0.0792.$$

This value is not very sensitive to the parameter r . It is natural to have a smaller value of α_c since only half of the neurons are used to store each pattern on average. It is noted that naively α_c is expected to be one half of that for the Hopfield model but the value obtained is not so small. This decrease of critical memory storage capacity cannot be serious for biological nervous systems like the brain since they consist of an enormous number of neurons.

As mentioned previously, F tends to be negative as q increases. In figure 2, F is shown, as an example, for $\alpha = 0.05$ when $q = 0.275$. The band of retrieval states almost disappears while that of the uncorrelated metastable states is almost unchanged. Thus patterns are stored in the present network by releasing spins at very small portions of the inhibitory neurons.

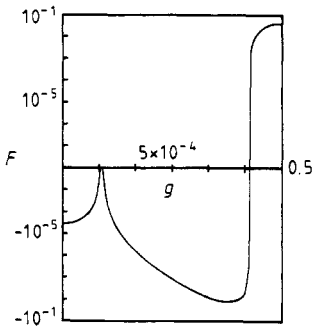


Figure 2. The function F for the physiological model with equal numbers of excitatory and inhibitory neurons when $\alpha = 0.05$ and $q = 0.275$.

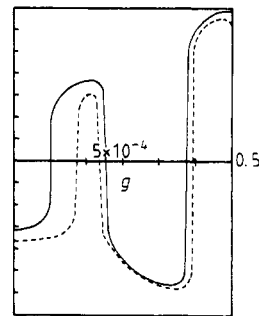


Figure 3. The function F for $\alpha = 0.06$ and $q = 0$ when the number of excitatory neurons and that of inhibitory neurons are not matched. The broken curve is for the excitatory-dominant case ($r = 0.8$) and the full curve is for the inhibitory-dominant case ($r = 0.2$).

Finally, we show some cases in which the numbers of excitatory and inhibitory neurons are not equal. Figure 3 shows the function F in two cases, $r = 0.2$ and 0.8 , when $\alpha = 0.06$ and $q = 0$. The two bands, especially the one corresponding to the retrieval states, are observed to grow as the number of inhibitory neurons increases. This seems to be natural, since, as mentioned above, most spins which are opposite to those of the original pattern are observed at the inhibitory neurons in the retrieval states. Therefore if the number of inhibitory neurons increases, the number of the retrieval states is expected to increase. However, this increase of the retrieval states does not necessarily lead to an improvement in the ability of memory retrieval in

inhibitory-dominant neural networks. Suppose that inhibitory neurons largely exceed excitatory neurons in number. Then the retrieval sequences easily enter a mode in which uniform configurations $\{1_i\}$ and $\{-1_i\}$ appear alternately. On the other hand, if excitatory neurons exceed inhibitory neurons, the sequences mostly end up in one of the uniform configurations. Thus, to keep the attraction basins of the retrieval states large enough compared with those for the uniform modes, the numbers of excitatory and inhibitory neurons are to be reasonably balanced. This point will be studied further in the next section.

4. The process of retrieval

To see the effect of incorporating the Dale hypothesis on the retrieval dynamics in some detail, we discuss statistical dynamics of overlaps (Amari 1977b, Kinzel 1985, Shinomoto 1987) for the present model. The above intuitive argument suggests that it is necessary to consider two kinds of overlaps: one is the overlap of the temporal pattern $\{S_i\}$ with the stored patterns $\{\xi_i^\mu\}$ and the other is that with the uniform mode $\{1_i\}$. Thus we introduce

$$m_\mu = \frac{1}{N} \sum_i S_i \xi_i^\mu \quad m_+ = \frac{1}{N} \sum_i S_i 1_i. \tag{8}$$

Next we define the referenced inputs u_i^μ and u_i^+ as follows:

$$u_i^\mu = \xi_i^\mu h_i \quad u_i^+ = 1_i h_i \tag{9}$$

with h_i given in (3). Then the overlaps at time $t + 1$ can be written in terms of the referenced inputs at time t as

$$m_\mu(t+1) = \frac{1}{N} \sum_i \text{sgn}(u_i^\mu(t)) \quad m_+(t+1) = \frac{1}{N} \sum_i \text{sgn}(u_i^+(t)). \tag{10}$$

The trick is to introduce an ensemble of a set of patterns satisfying

$$\langle \xi_i^\alpha \xi_j^\beta \rangle = \delta_{\alpha\beta} \delta_{ij} \quad \langle \xi_i^\alpha S_j \rangle = m_\mu \delta_{\mu\alpha} \delta_{ij} \quad \langle S_i \rangle = m_+ \quad \langle \xi_i^\alpha \rangle = m_{\mu+} \delta_{\mu\alpha} \tag{11}$$

where $\langle \dots \rangle$ stands for an average over the ensemble and $m_{\mu+}$ is the overlap between $\{\xi_i^\mu\}$ and $\{1_i\}$. In other words, we consider the ensemble over which the overlaps m_μ , m_+ and $m_{\mu+}$ take some fixed values. Note that the correlations among different sites and different patterns are neglected in the present approximation. Evaluating the mean values and the variances of u_i^μ and u_i^+ over the ensemble by using (11), we obtain

$$\bar{u}^\mu = \langle u_i^\mu \rangle = m_\mu \quad \bar{u}^+ = \langle u_i^+ \rangle = p(2r - 1)m_+ \tag{12}$$

and

$$\begin{aligned} \langle (u_i^\mu - \bar{u}^\mu)^2 \rangle &= \alpha + p^2 q^2 m_+^2 \\ \langle (u_i^+ - \bar{u}^+)^2 \rangle &= \alpha + m_\mu^2 \\ \langle (u_i^\mu - \bar{u}^\mu)(u_i^+ - \bar{u}^+) \rangle &= pqm_+ m_\mu \end{aligned} \tag{13}$$

where we have neglected the terms of $O(1/N)$.

Furthermore, when N goes to infinity, we can replace the RHS of (10) with the average of $\text{sgn}(u_i^\mu)$ or $\text{sgn}(u_i^+)$ over a two-dimensional Gaussian distribution $\rho(u^\mu, u^+)$

whose means and variances are given by (12) and (13):

$$\rho(\mathbf{u}^\mu, \mathbf{u}^+) = \left(\frac{\det M}{2\pi} \right)^{1/2} \exp\left[-\frac{1}{2}(\mathbf{u} - \mathbf{u}_0)^\top M(\mathbf{u} - \mathbf{u}_0)\right] \quad (14)$$

where

$$\mathbf{u} = (\mathbf{u}^\mu, \mathbf{u}^+)^\top \quad \mathbf{u}_0 = (\bar{u}^\mu, \bar{u}^+)^\top$$

$$M = \alpha^{-1}[\alpha + p^2(2r-1)^2 m_+^2 + m_\mu^2]^{-1} \begin{pmatrix} m_\mu^2 + \alpha & p(1-2r)m_\mu m_+ \\ p(1-2r)m_\mu m_+ & \alpha + p^2(1-2r)^2 m_+^2 \end{pmatrix}. \quad (15)$$

Introducing a vector notation to the overlaps in an obvious manner, we consider a nonlinear map \mathbf{G} from $\mathbf{m}(t)$ to $\mathbf{m}(t+2)$. Note that we consider the map for two time steps in order to take account of the uniform mode of period two appearing in inhibitory-dominant cases. Then it is a straightforward calculation to obtain the linearised map $\partial\mathbf{G}/\partial\mathbf{m}$ around a trivial fixed point $\mathbf{m} = 0$ of the map \mathbf{G} .

$$\left. \frac{\partial\mathbf{G}}{\partial\mathbf{m}} \right|_{\mathbf{m}=0} = \text{diag} \left(\frac{2}{\pi\alpha}, \frac{2p^2(1-2r)^2}{\pi\alpha} \right). \quad (16)$$

Now we find that the fixed point at the origin $\mathbf{m} = 0$ is stable if the conditions

$$\frac{2}{\pi\alpha} < 1 \quad \frac{2p^2(1-2r)^2}{\pi\alpha} < 1 \quad (17)$$

are satisfied. If one of the above inequalities is not satisfied, a couple of stable points appear on the corresponding instability axis through the pitchfork bifurcation and the origin turns into a saddle point. Since the bifurcation scenario with the present choice of the learning rule is essentially the same as the one with all-or-nothing learning, we do not repeat it here, but only compare the above stability conditions with those obtained for the all-or-nothing learning (Shinomoto 1987):

$$\frac{1}{\pi\alpha} < 1 \quad \frac{2p(1-2r)^2}{\pi\alpha} < 1.$$

We see that the origin stays unstable in the m_μ direction twice as long in the present model as α increases, which is preferable for the memory retrieval. This is natural since less information on the patterns is lost in the learning rule (1) than in the all-or-nothing learning. In short, the present learning rule teaches each pattern more firmly. On the contrary, the origin turns stable rather easily in the m_+ direction. This is interpreted as follows. In the all-or-nothing learning, half of the synaptic couplings T_{ij} are expected to vanish for any values of the parameter r if random patterns are learned. The present learning rule turns most of these vanishing couplings into negative or positive ones according to the inhibitory or excitatory nature of the corresponding neurons. Therefore the generated synaptic couplings develop an excitatory or inhibitory dominant nature more remarkably if the two types of neurons are not balanced in number. Thus the retrieval sequences comparatively easily enter into the uniform modes.

If the attraction basins of the uniform modes are desired to be reasonably small compared to those of the retrieval states, the second inequality of the stability conditions in (17) should be followed strongly. This requires a rather exact balance between

the numbers of excitatory neurons and inhibitory neurons. More precisely, the following inequality is to be obeyed:

$$\frac{1}{2} \left(1 - \frac{1}{p} \right) < r < \frac{1}{2} \left(1 + \frac{1}{p} \right). \quad (18)$$

Though the above argument is not exact, it is desirable that r is close to $\frac{1}{2}$ when p is large.

The inequality (18) follows only when we require rather wide attraction basins of the retrieval states. Therefore if input patterns which are not very close to the stored patterns are not to be identified with any one of them, r need not be $\sim \frac{1}{2}$. Then the unbalance of excitatory and inhibitory neurons in number makes it easier to drive retrieval sequences into the uniform modes from the input patterns not close to the stored patterns. This performance, which was called 'cognitive ability' (Shinomoto 1987), is a natural consequence of assigning unique excitatory or inhibitory character to each neuron.

5. Concluding remarks

We found that the number of the metastable states correlated to the stored patterns (i.e. the number of the retrieval states) increases remarkably at the cost of less critical storage capacity if each neuron has a unique excitatory or inhibitory character. In the retrieval states, most of spins opposite to those of the corresponding original pattern are observed at the inhibitory neurons. It will be interesting to study the possible implications of incorporating the Dale hypothesis in the retrieval dynamics of neural networks. The present choice of the learning is not sufficiently asymmetric to produce limit cycles in the retrieval dynamics. However, suppose that the degree of asymmetry is somehow increased. Then the dynamical structure of the network becomes fairly rich: the limit cycles will appear (Coolen and Ruijgrok 1988, Shiino *et al* 1989). It has also been argued that asymmetry may open chaotic trajectories (Parisi 1986). Then we may expect that there exist rich chaotic trajectories in the biological nervous systems since huge numbers of the metastable states around the stored patterns may provide a variety of branches for the retrieval trajectories. Actually we obtain asymmetric neural networks which exhibit chaotic retrieval dynamics under finite external (thermal) noises if the Dale hypothesis as well as time delay is incorporated (Fukai and Shiino 1989). The models and their retrieval dynamics will be reported elsewhere.

Acknowledgments

The author would like to thank Dr A Shudo for helpful discussions. He has also benefitted from discussions with Dr T Matsushita on various aspects of biological nervous systems.

Appendix

We show the derivation of (6) and (7) in the text. The starting equation is the RHS of

(5), which is written as

$$\prod_i \int dx_i \int_h \frac{d\tau_i}{2\pi} \left\langle \exp \left(i \sum_i x_i \tau_i + i(1-2g) \sum_{i=1}^N \varepsilon_i x_i - i\alpha \sum_i S_i x_i \sum_{j \neq i} S_j \eta_j - \frac{i}{N} \sum_{\mu \neq r} \sum_i \xi_i^\mu S_i x_i \sum_{j \neq i} \xi_j^\mu S_j \right) \right\rangle_\varepsilon \quad (A1)$$

where we have assumed, without any loss of generality, that the first Ng neurons are in opposite spin states to the pattern $\{\xi_i^r\}$, while the others are in the correct spin states: this implies that $\varepsilon_i = 1$ for $i = 1, \dots, Ng$ and -1 for $i = Ng + 1, \dots, N$.

Noting that

$$\sum_i S_i x_i \sum_{j \neq i} S_j \eta_j = \left(\sum_i S_i x_i \right) \left(\sum_j S_j \eta_j \right) - \sum_i \eta_i x_i \quad (A2)$$

$$\sum_{\mu \neq r} \sum_i \xi_i^\mu S_i x_i \sum_{j \neq i} \xi_j^\mu S_j = \sum_{\mu \neq r} \left(\sum_i \xi_i^\mu S_i x_i \right) \left(\sum_j \xi_j^\mu S_j \right) - (p-1) \sum_i x_i \quad (A3)$$

and making use of the results

$$\begin{aligned} \exp \left[-i\alpha \left(\sum_i S_i x_i \right) \left(\sum_j S_j \eta_j \right) \right] \\ = \int \frac{dz d\bar{z}}{2\pi\alpha} \exp \left(-\alpha^{-1} z\bar{z} + \bar{z} \sum_i S_i \eta_i - iz \sum_i S_i x_i \right) \end{aligned} \quad (A4)$$

$$\begin{aligned} \exp \left[\sum_{\mu \neq r} \left(\sum_i \xi_i^\mu S_i x_i \right) \left(\sum_j \xi_j^\mu S_j \right) \right] \\ = \prod_{\mu \neq r} \int \frac{dm_\mu d\bar{m}_\mu}{2\pi/N} \exp \left(-N \sum_{\mu \neq r} m_\mu \bar{m}_\mu + \sum_{\mu \neq r, i} (\bar{m}_\mu S_i \xi_i^\mu - im_\mu x_i S_i \xi_i^\mu) \right) \end{aligned} \quad (A5)$$

we can decouple each spin variable from the others. Now we take the average over patterns $\{\xi_i^\mu\}$ and the trace over $\{S_i\}$ to obtain

$$\begin{aligned} \prod_i \int dx_i \int_h \frac{d\tau_i}{2\pi} \int \frac{dm_\mu d\bar{m}_\mu}{2\pi/N} \int \frac{dz d\bar{z}}{2\pi\alpha} \\ \times \exp \left(i \sum_i x_i \tau_i + i(1-2g) \sum_{i=1}^N \varepsilon_i x_i + \sum_{\mu \neq r, i} \ln \cosh(\bar{m}_\mu - im_\mu x_i) \right. \\ \left. + \sum_i \ln \cosh(\bar{z}\eta_i - izx_i) + i\alpha \sum_i x_i (1 + \eta_i) \right. \\ \left. - N \sum_{\mu \neq r} m_\mu \bar{m}_\mu - \alpha^{-1} z\bar{z} - Ng \ln g - N(1-g) \ln(1-g) \right). \end{aligned} \quad (A6)$$

Expanding $\ln \cosh$, we obtain the Gaussian integrals for m_μ , \bar{m}_μ , z and \bar{z} . We find that the integrations of z and \bar{z} contribute only to the next leading order in N and thus are negligible. The integrations of m_μ and \bar{m}_μ are evaluated as

$$\begin{aligned} \prod_{\mu \neq r} \int \frac{dm_\mu d\bar{m}_\mu}{2\pi/N} \exp \left[\frac{1}{2} N \bar{m}_\mu^2 - \left(i \sum_i x_i + N \right) m_\mu \bar{m}_\mu - \frac{1}{2} \left(\sum_i x_i^2 \right) m_\mu^2 \right] \\ = \exp \left\{ -\frac{1}{2} \alpha N \ln \left[\frac{1}{N} \sum_i x_i^2 + \left(1 + \frac{i}{N} \sum_i x_i \right)^2 \right] \right\}. \end{aligned} \quad (A7)$$

After introducing new variables

$$A = \frac{1}{N} \sum_i x_i^2 \quad B = \frac{i}{N} \sum_i x_i \quad (\text{A8})$$

by using δ functions with their integral forms, we substitute (A7) into (A6). Then we perform the integrations of x_i and obtain

$$\int_h \prod_i \frac{d\tau_i}{\sqrt{2\pi}} \int \frac{da dA}{2\pi/N} \int \frac{db dB}{2\pi/N} \exp\left(-\frac{1}{4a} \sum_i [\tau_i + b + \alpha(1 + \eta_i) + (1 - 2g)\varepsilon_i]^2 + N\{-\frac{1}{2}\alpha \ln[A + (1 + B)^2] + aA - bB - g \ln g - (1 - g) \ln(1 - g)\}\right). \quad (\text{A9})$$

Now the integrations of A and B may be evaluated by the saddle point approximation. Defining the parameters r and q as given in the text, replacing the variables a and b as $\alpha/2a \rightarrow 1/a$ and $b + \alpha \rightarrow \beta\alpha$ and inserting γ in front of η_i in (A9), we finally obtain the desired equations (6) and (7).

References

- Amari S 1977a *Biol. Cybern.* **27** 77
 — 1977b *Biol. Cybern.* **26** 175
 Amit D J, Gutfreund H and Sompolinsky H 1985a *Phys. Rev. A* **32** 1007
 — 1985b *Phys. Rev. Lett.* **55** 1530
 — 1987 *Ann. Phys., NY* **173** 30
 Clark J W 1988 *Phys. Rep.* **158** 91
 Coolen A C C and Ruijgrok Th W 1988 *Phys. Rev. A* **38** 4253
 Eccles J C 1977 *The Understanding of the Brain* (New York: McGraw-Hill) 2nd edn
 Feigelman M V and Ioffe L B 1986 *Europhys. Lett.* **1** 197
 — 1987 *Int. J. Mod. Phys. B* **1** 51
 Fukai T and Shiino M 1989 *Preprint* Gumma Women's College/Tokyo Institute of Technology
 Gardner E 1986 *J. Phys. A: Math. Gen.* **19** L1047
 Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
 Kinzel W 1985 *Z. Phys. B* **60** 205
 Kohonen T 1977 *Associative Memory—A System Theoretical Approach* (Berlin: Springer) ch 4
 Parisi G 1986 *J. Phys. A: Math. Gen.* **19** L675
 Shiino M, Nishimori H and Ono M 1989 *J. Phys. Soc. Japan* **58** 763
 Shinomoto S 1987 *Biol. Cybern.* **57** 197
 Treves A and Amit D J 1988 *J. Phys. A: Math. Gen.* **21** 3155
 — 1989 *J. Phys. A: Math. Gen.* **22** 2205
 Virasoro M A 1988 *Europhys. Lett.* **7** 293