Mean-field theory for the Q-state Potts-glass neural network with biased patterns

1993 J. Phys. A: Math. Gen. 26 549

(http://iopscience.iop.org/0305-4470/26/3/017)

J. Phys. A: Math. Gen. 26 (1993) 549–562. Printed in the UK

Mean-field theory for the Q-state Potts-glass neural network with biased patterns

D Bollé†§, R Cools‡, P Dupont†|| and J Huyghebaert†

† Instituut voor Theoretische Fysica, Katolieke Universiteit Leuven, B-3001 Leuven, Belgium

‡ Departement Computerwetenschappen, Katolieke Universiteit Leuven, B-3001 Leuven, Belgium

Received 9 July 1992

Abstract. A systematic study of the Q-state Potts model of neural networks, extended to include biased patterns, is made for extensive loading α. Mean-field equations are written down within the replica symmetric approximation, for general Q and arbitrary temperature T. For the $Q = 3$ model and two classes of representative bias parameters, the storage capacity and retrieval quality at zero temperature are discussed as functions of the bias, taking into account the Mattis retrieval state and the lowest symmetric states. The T–α diagram is obtained and the stability properties of the retrieval state are analysed at finite temperatures. A comparison is made with the biased Hopfield model.

1. Introduction

Neural networks with multi-state neurons based upon Potts spins have attracted recent interest. This type of models has been first introduced by Kanter [1], where the storage capacity and retrieval of information were discussed (for symmetric synaptic couplings and unbiased stored patterns), mostly concentrating on the zero temperature limit. Since then this model has been further studied and extended in several directions taking into account e.g. more realistic neurophysiological properties of neurons. For a finite number of patterns the temporal development of the overlap has been studied for asymmetric couplings, allowing non-zero correlations (≡ bias) between the patterns [2–4]. Also the stability of these networks at zero and finite temperatures has been extensively treated as a function of bias [5]. Later it has been shown that this type of models (the Potts perceptron) can be used successfully in multiclass classification problems [6, 7] and in a statistical formulation of the Willshaw model with local inhibition [8]. Very recently local inhibition has also been applied in layered feedforward networks with multi-state elements leading to a Potts-glass version of these networks [9]. Furthermore an isotropic version of the model has been shown to be particularly suited for invariant pattern recognition [10].

As for largely mathematical approaches, the parallel dynamics of a diluted version of the model has been solved exactly using a probabilistic approach [11, 12]. A

§ E-mail address: FGBDA18@BLEKUL11.BITNET.

|| Present address: Department of Nuclear Medicine, University Hospital Gasthuisberg, B-3000 Leuven, Belgium.

0305-4470/93/030549+14$07.50

rigorous lower bound for the storage capacity of Kanter's Potts model has been derived in [13] and the almost sure convergence of the free energy and the overlap order parameter in the thermodynamic limit have been proved in [14] for a number of stored patterns growing logarithmically with the size of the system. A systematic study of the Q-state Potts glass neural network with extensive loading of biased patterns is quite tedious and, to our knowledge, has not been performed so far. The purpose of the present paper is to provide such a study. Preliminary results on the thermodynamic properties of these networks, mostly concentrating on unbiased patterns, have been given in [15].

This paper is organized as follows. In section 2 we describe the model in detail. In section 3 we write the free energy and the definition of the order parameters in mean-field theory. Within the replica-symmetric approximation, we find the fixed-point equations at arbitrary temperatures for these order parameters. In the limit of zero temperature these equations can be reformulated into two coupled equations for the appropiate combinations of order parameters. We then discuss in detail in section 4 the solutions of these equations for the retrieval states and the lowest-order symmetric states of the $Q = 3$ model with two representative classes of bias parameters. In particular the critical storage capacity and the quality of retrieval are analysed and compared with the biased Hopfield model. Section 5 presents the finite temperature results for the retrieval states of the same models. A temperature-capacity diagram is given as a function of the bias parameters and the stability region for the retrieval states is obtained as a function of temperature, bias and storage capacity. Our findings are compared with similar results for the biased Hopfield model. In section 6 the conclusions are presented. Finally, the fixed-point equations for the $Q = 3$ model at zero temperature are written out in detail in the appendix.

2. The model

Consider a system of N neurons. Each neuron can be described by a Potts spin $\sigma_i \in \{1,2,\ldots,Q\}$, $i = 1,2,\ldots,N$. The neurons are interconnected with all the others by a synaptic matrix of strength J_{ij}^{kl} which determines the contributions of a signal fired by the jth presynaptic neuron in state l to the postsynaptic potential which acts on the ith neuron in state k. The energy potential h_{i,σ_i} of a neuron i which is in a state σ_i is given by

$$h_{i,\sigma_i} = -\sum_{\substack{j=1 \\ j\neq i}}^{N} \sum_{k,l=1}^{Q} J_{ij}^{kl} u_{\sigma_i,k} u_{\sigma_j,l} \tag{1}$$

with the Potts spin operator u defined as

$$u_{\sigma_i,k} = Q\delta_{\sigma_i,k} - 1. \tag{2}$$

We assume that the synaptic couplings are symmetric, i.e. $J_{ij}^{kl} = J_{ji}^{lk}$. The dynamics of the Q-state Potts model is defined as in [1]. At zero temperature the state of the neuron in the next time step is fixed to be the state which minimizes the induced local field (1). The stable states of the system are those configurations $\{\sigma_i\}$ where every neuron is in a state which gives a minimum value to $\{h_{i,\sigma_i}\}$. For symmetric

couplings this stability is equivalent to the requirement that the configurations $\{\sigma_i\}$ are the local minima of the Potts Hamiltonian

$$H = -\frac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\sum_{k,l=1}^{Q} J_{ij}^{kl} u_{\sigma_i,k} u_{\sigma_j,l}. \tag{3}$$

In the presence of noise there is a finite probability of having configurations other than the local minima. This can be taken into account by introducing an effective temperature $T = 1/\beta$.

To build in the capacity for learning and memory in this network, its stationary configurations representing the retrieved patterns must be correlated with the stored patterns $\{\xi^\mu\}$, $\mu = 1, 2..., p$ fixed by the learning process. The latter are allowed to be correlated, i.e. the $\{\xi^\mu\}$ are chosen as independent random variables which can take the values $1, 2..., Q$ with probability

$$P(\xi_i^\mu = k) = \frac{1 + B_k}{Q}, \quad k = 1, 2..., Q \tag{4}$$

where $\{B_k\}$ are the bias parameters. Since $P(k)$ are probabilities, $\{B_k\}$ satisfy the relations

$$-1 \leqslant B_k \leqslant Q - 1 \qquad \sum_{k=1}^{Q} B_k = 0. \tag{5}$$

As in [5], we therefore propose the learning rule

$$J_{ij}^{kl} = \frac{1}{Q^2 N}\sum_{\mu=1}^{p}\left(u_{\xi_i^\mu,k} - B_k\right)\left(u_{\xi_j^\mu,l} - B_l\right). \tag{6}$$

In what follows, we study this biased Potts network for finite loading $\alpha = p/N$, $p \to \infty$, $N \to \infty$. At this point we remark that the $\alpha = 0$ results for this model have been studied in [5], the $B_k = 0, k = 1, 2..., Q$ case for finite α has been discussed in [1] for a learning rule of the type (6) and in [10] for the learning rule

$$J_{ij}^{kl} = \frac{1}{Q^2 N} u_{k,l} \sum_{\mu=1}^{p} u_{\xi_i^\mu,\xi_j^\mu} \tag{7}$$

which stores only the information whether two neurons are in the same state or not.

3. Replica-symmetric mean field theory

The Hamiltonian (3) for the learning rule (6) can be rewritten as

$$H = -\frac{1}{2N}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\sum_{\mu=1}^{p}\left(u_{\xi_i^\mu,\sigma_i} - B_{\sigma_i}\right)\left(u_{\xi_j^\mu,\sigma_j} - B_{\sigma_j}\right) \tag{8}$$

where we have employed the identity

$$\frac{1}{Q}\sum_{k=1}^{Q} u_{\sigma_i,k}u_{\sigma_j,k} = u_{\sigma_i,\sigma_j}. \tag{9}$$

Using standard techniques [1, 16], one finds the following expression for the free energy density in the limit $N \to \infty$ and for s condensed patterns $\{\xi_i^\nu\}, \nu = 1,2,\ldots,s$:

$$f = \lim_{n\to 0} f_n,$$

$$f_n = \frac{1}{2n}\sum_{\lambda=1}^{n}\sum_{\nu=1}^{s}(m_\nu^\lambda)^2 + \frac{\alpha}{2n\beta}\operatorname{Tr}\ln[\mathbb{1} - \beta q] + \frac{\alpha\beta}{n}\sum_{\substack{\rho,\lambda=1\\ \lambda<\rho}}^{n} r_{\lambda\rho}q_{\lambda\rho} + \frac{\alpha}{2n}\sum_{\lambda=1}^{n} q_{\lambda\lambda}$$

$$-\frac{1}{n\beta}\left\langle\!\!\left\langle \ln \operatorname{Tr}_{\sigma^\lambda} \exp\left[\beta\sum_{\lambda=1}^{n}\sum_{\nu=1}^{s}(m_\nu^\lambda + h_\nu)(u_{\xi^\nu,\sigma^\lambda} - B_{\sigma^\lambda}) + \alpha\beta^2\sum_{\substack{\rho,\lambda=1\\ \lambda<\rho}}^{n}\sum_{k=1}^{Q}\left[\frac{1+B_k}{Q}\right](u_{k,\sigma^\lambda} - B_{\sigma^\lambda})(u_{k,\sigma^\rho} - B_{\sigma^\rho})\right]\right\rangle\!\!\right\rangle \tag{10}$$

with h_ν the couplings of the external field terms and with $\langle\!\langle \cdots \rangle\!\rangle$ denoting the quenched average over the distribution of the condensed patterns $\{\xi_i^\nu\}$. The order parameters $m_\nu^\lambda, q_{\lambda\rho}$ and $r_{\lambda\rho}$ have the following meaning

$$m_\nu^\lambda = \frac{1}{N}\sum_{i=1}^{N}\left\langle\!\!\left\langle < u_{\xi_i^\nu,\sigma_i^\lambda} - B_{\sigma_i^\lambda} > \right\rangle\!\!\right\rangle \tag{11}$$

$$q_{\lambda\rho} = \frac{1}{N}\sum_{i=1}^{N}\left\langle\!\!\left\langle \sum_{k=1}^{Q}\left[\frac{1+B_k}{Q}\right] < u_{k,\sigma_i^\lambda} - B_{\sigma_i^\lambda} >< u_{k,\sigma_i^\rho} - B_{\sigma_i^\rho} > \right\rangle\!\!\right\rangle \tag{12}$$

$$r_{\lambda\rho} = \frac{1}{\alpha}\sum_{\mu=s+1}^{p}\langle\!\langle m_\mu^\lambda m_\mu^\rho \rangle\!\rangle \tag{13}$$

where $< \cdots >$ stands for thermal average. Hence, m_ν^λ describe the macroscopic overlap with the condensed patterns, $q_{\lambda\rho}$ are the extended Edwards–Anderson order parameters and the $r_{\lambda\rho}$ describe the overlap with the noncondensed patterns. We note that in contrast with the Hopfield model with biased patterns [17], the order parameters $q_{\lambda\rho}$ explicitly contain the bias. Consistently we get that for $Q = 2$, using $u_{\sigma_i,\sigma_j} = \sigma_i\sigma_j$ and $B_\sigma = \sigma a$, where the σ are now the Ising spins ± 1, equation (12) reduces to

$$q_{\lambda\rho} = (1-a^2)\frac{1}{N}\sum_{i=1}^{N}\left\langle\!\!\left\langle < \sigma_i^\lambda >< \sigma_j^\rho > \right\rangle\!\!\right\rangle \tag{14}$$

such that in that case the multiplicative bias factor $(1-a^2)$ can be taken out of the definition of $q_{\lambda\rho}$.

Assuming the replica symmetry, i.e.

$$\begin{aligned} &m_\nu^\lambda = m_\nu \\ &q_{\lambda\rho} = q \qquad r_{\lambda\rho} = r \qquad \text{for } \lambda \neq \rho \qquad \lambda, \rho = 1, 2, \ldots, n \\ &q_{\lambda\rho} = \tilde{q} \qquad r_{\lambda\rho} = \tilde{r} \qquad \text{for } \lambda = \rho \end{aligned} \tag{15}$$

where the diagonal values of $q_{\lambda\rho}$ and $r_{\lambda\rho}$ have to be introduced as separate order parameters since they are functions of σ_i (in contrast with the biased Hopfield model [17] and with the anisotropic Potts model [1]), in the limit $n \to 0$ the free energy density (10) becomes

$$\begin{aligned} f = \frac{1}{2}\sum_{\nu=1}^{s} m_\nu^2 - \frac{1}{2}\alpha\beta r q + \frac{1}{2}\alpha\tilde{q} - \frac{\alpha q}{2[1-\beta(\tilde{q}-q)]} + \frac{\alpha}{2\beta}\ln[1-\beta(\tilde{q}-q)] \\ + \frac{1}{2}\alpha\beta\tilde{r}\tilde{q} - \frac{1}{\beta}\left\langle\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \ln \sum_{\sigma=1}^{Q} \exp\left[\beta\mathcal{H}_\sigma(\xi, z)\right] \right\rangle\!\right\rangle . \end{aligned} \tag{16}$$

with the Gaussian measure $\mathrm{D}z$ given by

$$\mathrm{D}z = \prod_{k=1}^{Q} dz_k (2\pi)^{-1/2} \exp(-z_k^2/2). \tag{17}$$

Furthermore $\mathcal{H}_\sigma(\xi, z)$ reads

$$\begin{aligned} \mathcal{H}_\sigma(\xi, z) = \tfrac{1}{2}\alpha\beta(\tilde{r}-r)(1+B_\sigma)(Q-1-B_\sigma) + \sum_{\nu=1}^{s}(u_{\xi^\nu,\sigma} - B_\sigma)(m_\nu + h_\nu) \\ + \sum_{k=1}^{Q}\sqrt{\alpha r(1+B_k)/Q}\,(u_{k,\sigma} - B_\sigma) z_k . \end{aligned} \tag{18}$$

From (16)–(18) we get the fixed-point equations for the order parameters

$$m_\nu = \left\langle\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \frac{\sum_\sigma (u_{\xi^\nu,\sigma} - B_\sigma)\exp(\beta\mathcal{H}_\sigma(\xi, z))}{\sum_\sigma \exp(\beta\mathcal{H}_\sigma(\xi, z))} \right\rangle\!\right\rangle \tag{19}$$

$$q = \tilde{q} - \frac{1}{\beta}\sum_{k=1}^{Q}\sqrt{\frac{1+B_k}{\alpha r Q}}\left\langle\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \frac{\sum_\sigma (u_{k,\sigma} - B_\sigma) z_k \exp(\beta\mathcal{H}_\sigma(\xi, z))}{\sum_\sigma \exp(\beta\mathcal{H}_\sigma(\xi, z))} \right\rangle\!\right\rangle \tag{20}$$

$$\tilde{q} = \left\langle\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \frac{\sum_\sigma (Q-1-B_\sigma)(1+B_\sigma)\exp(\beta\mathcal{H}_\sigma(\xi, z))}{\sum_\sigma \exp(\beta\mathcal{H}_\sigma(\xi, z))} \right\rangle\!\right\rangle \tag{21}$$

$$r = q[1-\beta(\tilde{q}-q)]^{-2} \tag{22}$$

$$\tilde{r} = -\beta^{-1} + [2q - \tilde{q} + \beta^{-1}][1-\beta(\tilde{q}-q)]^{-2}. \tag{23}$$

We note that in general two of the order parameters, i.e. r and $\tilde{r}$ are algebraic. Furthermore for zero bias, $\tilde{q} = Q - 1$.

In the limit of zero temperature the fixed-point equations (19)–(23) can be rewritten in the following way. Taking for simplicity $s = 1$, i.e. only one condensed pattern, we first rewrite

$$\beta(\tilde{q} - q) = C \tag{24}$$

$$\beta(\tilde{r} - r) = C(1 - C)^{-1} \tag{25}$$

with

$$C = \frac{1}{\sqrt{\alpha r Q}} \left\langle\!\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \sum_{\sigma=1}^{Q} \left[\left(Q\sqrt{1+B_\sigma}\, z_\sigma - B_\sigma \sum_{k=1}^{Q} \sqrt{1+B_k}\, z_k \right) \right.\right.\right.$$

$$\left.\left.\left. \times (Z_\sigma(\xi, z))^{-1} \right] \right\rangle\!\!\right\rangle \tag{26}$$

$$Z_\sigma(\xi, z) = 1 + \sum_{\substack{\rho=1 \\ \rho \neq \sigma}}^{Q} \exp \beta(\mathcal{H}_\rho(\xi, z) - \mathcal{H}_\sigma(\xi, z)) \tag{27}$$

and $\mathcal{H}(\xi, z)$ given by (18). At this point we note that the order parameters m and $\tilde{q}$ and hence r and $\tilde{r}$ can also be written in a form where the β-dependence is only located in the function Z.

In the limit $\beta \to \infty$ the sign of $(\mathcal{H}_\rho - \mathcal{H}_\sigma) \equiv \Delta_{\rho\sigma}$ determines the important fact whether $(Z_\sigma)^{-1}$ is 0 or 1. Consequently the system can be described by the solutions of

$$m = -1 + Q \left\langle\!\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \prod_{\substack{\rho=1 \\ \rho \neq \xi}}^{Q} \Theta(-\Delta_{\rho\xi}(\xi, z)) \right\rangle\!\!\right\rangle$$

$$- \sum_{\sigma=1}^{Q} B_\sigma \left\langle\!\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \prod_{\substack{\rho=1 \\ \rho \neq \sigma}}^{Q} \Theta(-\Delta_{\rho\sigma}(\xi, z)) \right\rangle\!\!\right\rangle \tag{28}$$

$$C = \frac{1}{\sqrt{\alpha r Q}} \left\langle\!\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \sum_{\sigma=1}^{Q} \left(Q\sqrt{1+B_\sigma}\, z_\sigma - B_\sigma \sum_{k=1}^{Q} \sqrt{1+B_k}\, z_k \right) \right.\right.$$

$$\left.\left. \times \prod_{\substack{\rho=1 \\ \rho \neq \sigma}}^{Q} \Theta(-\Delta_{\rho\sigma}(\xi, z)) \right\rangle\!\!\right\rangle \tag{29}$$

$$r = \frac{1}{(1-C)^2} \sum_{\sigma=1}^{Q} (Q - 1 - B_\sigma)(1 + B_\sigma) \left\langle\!\!\left\langle \int_{\mathbb{R}^Q} \mathrm{D}z \prod_{\substack{\rho=1 \\ \rho \neq \sigma}}^{Q} \Theta(-\Delta_{\rho\sigma}(\xi, z)) \right\rangle\!\!\right\rangle . \tag{30}$$

To get a contribution from the Θ function in (28)–(30) we need to satisfy $Q - 1$ conditions $\Delta_{\rho\sigma}(\xi, z) < 0$ for $\rho = 1, 2, \ldots, Q, \rho \neq \sigma$ i.e.

$$\sum_{k=1}^{Q} \sqrt{1+B_k}(u_{k,\rho} - u_{k,\sigma} - (B_\rho - B_\sigma)) z_k$$

$$< -\frac{\alpha}{2}\frac{C}{1-C}\sqrt{\frac{Q}{\alpha r}}(B_\rho - B_\sigma)(Q-2-B_\rho - B_\sigma)$$
$$- m\sqrt{\frac{Q}{\alpha r}}(u_{\xi,\rho} - u_{\xi,\sigma} - (B_\rho - B_\sigma)). \tag{31}$$

Looking closely at these conditions one realizes that the following two combinations of order parameters are important

$$x = \frac{C}{1-C}\sqrt{\frac{Q}{2\alpha r}} \qquad y = m\sqrt{\frac{Q}{2\alpha r}}. \tag{32}$$

This implies that the equations for the order parameters for the Q-state Potts model with biased patterns can be transformed into two coupled equations for x and y. For zero bias we get a further reduction to one single nonlinear equation for y since the conditions (31) simplify to ($\rho = 1,2,\ldots,Q, \rho \neq \sigma$)

$$z_\rho < z_\sigma + \frac{\sqrt{2}}{Q}(u_{\xi,\rho} - u_{\xi,\sigma})y. \tag{33}$$

This equation for y is identical to the scaled equation written down in [10].

4. Zero temperature results for the $Q = 3$ model

In this section we start by rewriting the fixed-point equations (28)–(30) for $Q = 3$ and n-symmetric states $m = m_n(1,1,\ldots,1,0,0,\ldots,0)$ where the first n components are unity and the remaining components are zero. The retrieval states are given by $n = 1$.

To simplify and speed up the numerical treatment of these equations we make a transformation of the integration variables z such that one of the integrals is trivially given by zero or one, and in fact only two-dimensional integrals survive. For $T = 0$ they can be further reduced to one-dimensional integrals. This is explained in detail in the appendix.

These equations (A9)–(A16) are then solved numerically as a function of the bias parameters $\boldsymbol{B}$. Let us first remark that it is convenient to rewrite the bias $\boldsymbol{B} = (B_1, B_2, \ldots, B_Q)$ in a different form, namely

$$\boldsymbol{B} = a(b_1, b_2, \ldots, b_Q) \qquad b_1 \geqslant b_2 \ldots \geqslant b_Q \qquad a \in [0,1]. \tag{34}$$

We call a the bias amplitude and $b_1, b_2, \ldots, b_Q$ the bias structure. Due to the fact that the model described in section 2 is invariant under permutations of a neuron i, (34) is not an additional assumption.

We have then selected two representative classes of bias parameters, i.e. $\boldsymbol{B}_1 = a(2,-1,-1)$ and $\boldsymbol{B}_2 = a(1,0,-1)$. The form $\boldsymbol{B}_1$ indicates that one state is privileged and the other states have equal probability to appear. In fact, for $a = 1$ the probabilty distribution for the patterns is such that the lowest state has probability one. This means that there is no freedom left for the neurons. In the other case, $\boldsymbol{B}_2$, all three states have different probability. The distribution for the patterns is such that two states have non-zero probability. Hence the neurons can still occupy different states.

Before discussing the numerical results we write down the analytic expressions for the free energy and the entropy. From (16) we find in the limit $\beta \to \infty$ after a straightforward calculation that

$$f = -\frac{1}{2}\sum_{\nu=1}^{s}(m_\nu)^2 - \frac{1}{2}\alpha r C(2-C) \tag{35}$$

and that the entropy S is given by

$$S = \beta^2 \frac{\partial f}{\partial \beta} = -\frac{1}{2}\alpha\left[\ln(1-C) + \frac{C}{1-C}\right] \tag{36}.$$

These expressions are formally identical to their analogues in the Potts model without bias and the Hopfield model.

Let us then turn to the explicit results. The numerical analysis of (A9)–(A16) has been performed using the techniques described in [18]. In figure 1 the critical storage capacity is shown for the positive Mattis state ($m > 0$) as a function of the bias amplitude a. Since in the definition of the order parameter m_ν the overlap due to the non zero correlations between the patterns is subtracted, it is clear that $\alpha_c(a=1) = 0$ for $\boldsymbol{B} = \boldsymbol{B}_1$. This is similar to the Hopfield model [17]. Furthermore we recall that $\alpha_c = 0.415$ for $a = 0$ [1]. Since $\boldsymbol{B}_1$ is the most extreme choice of bias structure possible [5], all $Q = 3$ systems have a Mattis storage capacity lying between the curve $\boldsymbol{B}_1$ and the line $\alpha_c = 0.415$.

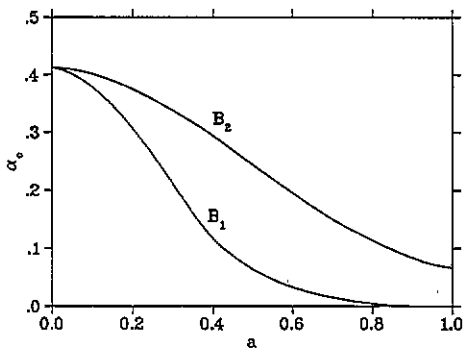

Figure 1. The critical capacity α_c as a function of the bias amplitude a for the $Q = 3$ $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ Potts networks at $T = 0$.

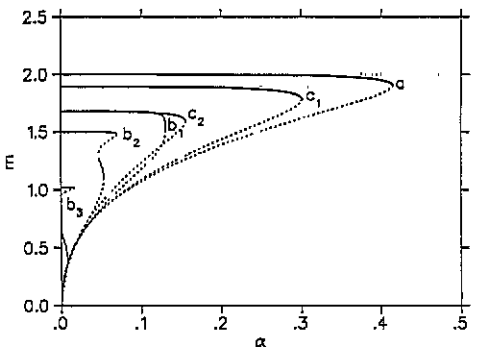

Figure 2. The overlap m for the retrieval state as a function of α at $T = 0$: curve (a) $\boldsymbol{B} = 0$; curve (b_1) $\boldsymbol{B}_1$ model for $a = 0.4$; curve (b_2) $\boldsymbol{B}_1$ model for $a = 0.5$; curve (b_3) $\boldsymbol{B}_1$ model for $a = 0.7$; curve (c_1) $\boldsymbol{B}_2$ model for $a = 0.4$; curve (c_2) $\boldsymbol{B}_2$ model for $a = 0.7$. The full (dotted) lines represent stable (unstable) solutions.

The corresponding overlap is given in figure 2. For zero bias the overlap is nearly maximal over a long interval in α, indicating almost perfect retrieval properties of the system. Increasing the bias amplitude we see that for the $\boldsymbol{B}_2$ system the bifurcation diagram retains the same form but the overlap is decreasing; for $a = 1$ we are left with the point $m = 4/3$ because from [5] we know that for $\alpha = 0$

$$m = Q - 1 - \frac{a^2}{Q}\sum_{k=1}^{Q} b_k^2. \tag{37}$$

For the B_1 system the overlap decreases more quickly in function of the bias amplitude until it reaches $m = 0$ for $a = 1$ (see again (37)). We also remark that from $a = 0.4$ onwards until $a = 1$ the bifurcation diagram shows a second turning point. This turning point always appears at $\alpha < \alpha_c$. On these figures also the stability of the solutions is indicated. The stability results were obtained in a standard way [19] by studying numerically the eigenvalues of the Hessian matrix formed by the second derivatives of the free energy with respect to the order parameters given in (15).

For the symmetric states we have restricted ourselves to $n = 2$ and $n = 3$. In figure 3 (full lines) the storage capacity is shown for the B_1 model as a function of a and figure 4 (full lines) does the same for the B_2 model. For the B_1 model we see that there is a pronounced peak and a higher capacity than for the B_2 model, except for a close to 1. The capacity for the symmetric states is larger than that for the Mattis state for $a > 0.411$ ($n = 2$) and $a > 0.551$ ($n = 3$). For comparison we show in figure 3 (dotted curves) the results for the biased Hopfield model. Here we note the presence of the second peak around $a = 0.7$ for the $n = 3$ symmetric state that is not explicitly shown in [17]. Also here the capacity for the symmetric states is larger than that for the retrieval state for values of the bias $a > 0.394$ ($n = 2$) and $a > 0.552$ ($n = 3$). For the B_2 model the storage capacity for the $n = 2, 3$ states is roughly a factor of 3 lower than that for the $n = 1$ retrieval state.

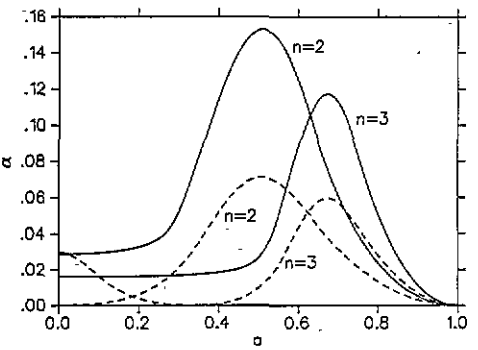

Figure 3. The storage capacity α as a function of a for the $n = 2$ and $n = 3$ symmetric states in the $Q = 3$ B_1 model (full curves) and the Hopfield model (dashed curves) at $T = 0$.

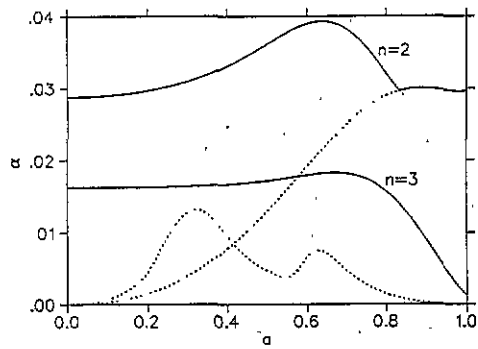

Figure 4. As figure 3, but for the $Q = 3$ B_2 model (full curves). The dotted curves represent the turning point at the smallest capacity (see figure 5).

The dotted curves in figure 4 represent, for the B_2-network, the turning point at the smallest capacity. This can be clearly understood from figure 5 where, as an illustration, the overlap is shown for the $n = 2$ symmetric state for both the B_1 and B_2 system. Indeed, we see the appearance of two turning points, especially for the B_2 system (figure 5(b)), for certain bias regions. We note that in contrast with the B_1 system the turning point with the highest overlap occurs at a loading $\alpha < \alpha_c$ for $a < 0.83$. In the case of the $n = 3$ symmetric state we even find the possibility of three turning points.

To complete the discussion of the $T = 0$ behaviour of the network we mention two further results. Firstly, we know that for the biased Hopfield system the Mattis states are no longer the ground states when the bias parameter a exceeds the value $a = (\sqrt{2} - 1)$ [17]. For the $Q = 3$ Potts systems with bias, results of this type are

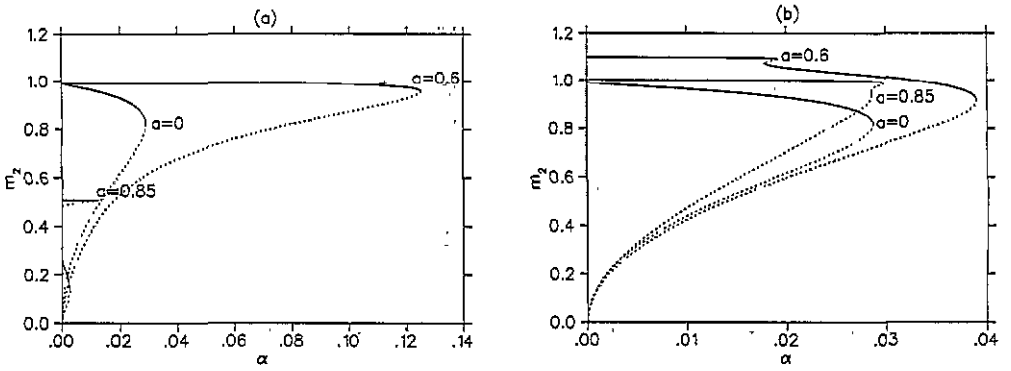

Figure 5. The overlap m_2 for the $n = 2$ symmetric states as a function of α for (*a*) the $Q = 3$ B_1 model and (*b*) the B_2 model at $T = 0$ for different values of a. The full (dotted) lines represent stable (unstable) solutions.

strongly dependent on the specific bias parameters. We have already seen in [5] that in the case of $\alpha = 0$ the Mattis state and the first symmetric state for the B_1 system cross when the bias amplitude is bigger than $(\frac{1}{8}\sqrt{18\sqrt{2}-5}+3\sqrt{2}-5)$ while for the B_2 system the Mattis state remains the lowest in energy for all values of the bias amplitude. This behaviour persists if $\alpha \neq 0$.

Secondly, to have an idea about the internal consistency of the replica-symmetric theory we have calculated the entropy at $T = 0$ using expression (36). For the ferromagnetic phase we know already that the Hopfield model without bias leads to an entropy at the critical capacity equal to $S_{\alpha_c} = -1.4 \times 10^{-3}$. For the $Q = 3$ Potts model without bias we find -3.3×10^{-3}. For the B_1 system we find for $a = 0.4$ the value -1.5×10^{-2}, for $a = 0.7$ the value -8.8×10^{-5}. For the B_2 system the entropy is -5.9×10^{-3} for $a = 0.4$ and -4.6×10^{-3} for $a = 0.7$. All this suggests that replica symmetry breaking is still weak in the retrieval states for the $Q = 3$ Potts model with bias.

5. Results at finite temperature

An analysis of these models at a finite temperature is much more involved. In the following we restrict ourselves to the positive Mattis states which are most important to determine the retrieval quality of the network. In particular, we do not discuss the transition from the disordered paramagnetic phase to the spin-glass phase. For results about these thermodynamic properties of the model we refer to [15].

The fixed point equations (19)–(23) now contain two-dimensional integrals as explained at the beginning of section 4. These integrals are calculated using the specific numerical techniques described in [20, 21]. Employing a bifurcation analysis we arrive at the phase diagram depicted in figures 6 and 7 for the B_1 and B_2 system, respectively. When crossing the line T_M from above Mattis retrieval states show up as local minima of the free energy. At this point the overlap with the built-in patterns jumps from zero to a finite macroscopic value. So the system functions as an associative memory and the critical storage capacity for a given temperature can be read off through the line T_M. When T is lowered further, the retrieval states become global minima of the free energy [15]. This happens along the line T_c. The

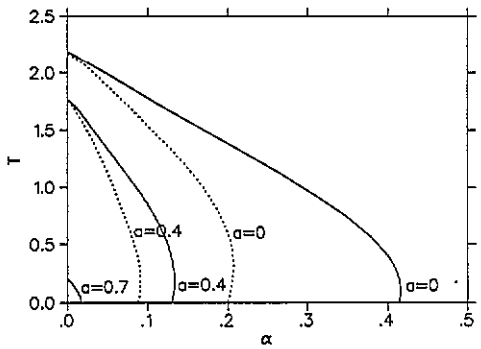

Figure 6. A (T, α) phase diagram for different values of a for the $Q = 3$ B_1 model. The full lines denotes T_M, the dotted lines denotes T_c.

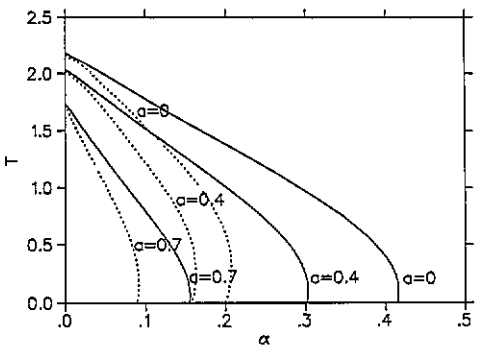

Figure 7. As figure 6, but for the $Q = 3$ B_2 model.

transition at T_c is a first-order transition. We note that for the B_1 model, the line T_c for $a = 0.7$ is not visible on the scale of figure 6.

In contrast with the Hopfield model the critical lines T_M and T_C are different for $\alpha = 0$ due to the fact that even for a finite number of patterns the Potts model has a discontinuous transition at T_c [5]. Furthermore, in all cases shown here we find weak reentrant spin-glass behaviour, analogous to what has been seen [22] in the Hopfield model.

As expected, there is not much difference in shape for the critical lines when introducing a non-zero bias. The value of the critical capacity for the B_1 system decreases more quickly as a function of the bias amplitude a than that for the B_2 system. For zero bias we confirm the known results $\alpha_c = 0.415$ [1] and $T_M(\alpha = 0) = 2.185$ [2, 10], $T_c(\alpha = 0) = 2.165$ [5].

It is interesting to know the stability region of the retrieval states as a function of T and a over the whole range of these parameters. These regions are depicted in figures 8 and 9 for B_1 and B_2, respectively, at two values of α and compared with their $\alpha = 0$ analogues. For the B_1 system (figure 8), the following remarks can be made. The disconnected stability region in the $\alpha = 0$ case in the interval $a \in [0.6, 1]$ disappears for $\alpha \neq 0$. For example, for $\alpha = \alpha_c(T = 0, a = 0)/10$ the 'knee' around $a = 0.55$ is still reminiscent for this disconnectedness while for $\alpha = \alpha_c(T = 0, a = 0)/2$ the behaviour is smooth. For the B_2 system (figure 9), no such disconnectedness appears in the stability region. Also, for $a = 1$ the temperature for $0 \leqslant \alpha \leqslant 6.44 \times 10^{-2}$ is non-zero in contrast with the B_1 system due to the fact mentioned in section 4 that for the B_2 system all three states have different probability. For completeness we present similar results for the Hopfield model in figure 10. The $Q = 3$ B_1 system shows some resemblance with the Hopfield model.

6. Concluding remarks

In this paper we have completed the study of the Potts model with bias, started in [5] for low loading $\alpha = 0$, by treating the case of finite loading $\alpha \neq 0$. Fixed-point equations at finite temperature are derived for general Q and in the limit of zero temperature it is shown how they reduce to two coupled equations for appropiate combinations of the order parameters. The $Q = 3$ system has been solved explicitly at

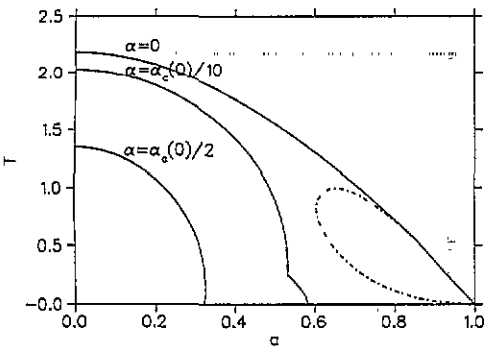

Figure 8. The temperature T_M as a function of a for the $Q = 3$ B_1 model for different values of α. The chain curve indicates the border of the stability region.

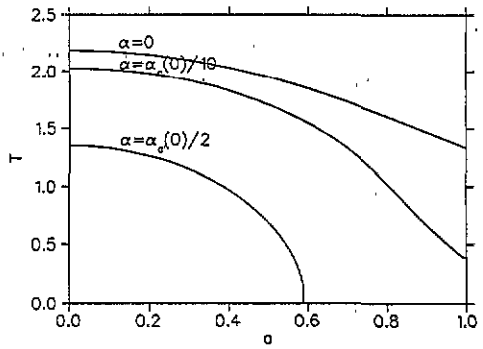

Figure 9. As figure 8, but for the $Q = 3$ B_2 model.

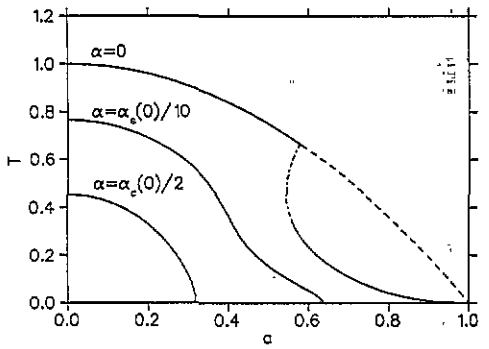

Figure 10. As figure 8, but for the Hopfield model.

finite loading α for two representative classes of bias parameters $B_1 = a(2,-1,-1)$ and $B_2 = a(1,0,-1)$. In particular, the positive Mattis retrieval state and the lowest symmetric states have been discussed. At zero temperature, we find that the Mattis storage capacity is higher for the B_2 system than for the B_1 system over the whole a interval while the storage capacity for the $n = 2$ and $n = 3$ symmetric states has the inverse behaviour. The Mattis overlap for B_1 decreases more quickly than the overlap for B_2 as a function of a.

At non-zero temperatures, α_c for the B_1 system also decreases more quickly. For $\alpha = 0$ a disconnected stability region for the Mattis state exists. In all its properties the B_1 model shows some resemblance with the Hopfield model.

Acknowledgments

This work has been supported in part by the Research Fund of K U Leuven (grant OT/91/13). The authors are indebted to R Kühn, L van Hemmen and B Vinck for stimulating discussions. D B and J H would like to thank the Belgian National Fund for Scientific Research and the Inter-University Institute for Nuclear Sciences for financial support.

Appendix

The reformulation of the zero temperature fixed-point equations (28)–(30) for the $Q = 3$ n-symmetric states is based upon the following transformation

$$y = Tz \tag{A1}$$

$$T_{k,1} = \sqrt{\frac{1+B_k}{3}} \qquad k = 1,2,3 \tag{A2}$$

$$T_{k,2} = \frac{1}{N_2}\sqrt{\frac{\alpha}{3}(1+B_k)}\,(U_{k,2} - U_s U_{k,3}) \tag{A3}$$

$$T_{k,3} = \frac{1}{N_3}\sqrt{\frac{\alpha}{3}(1+B_k)}\;U_{k,3} \tag{A4}$$

where the Us involve Potts operators and the N_2 and N_3 are normalisation constants respectively given by

$$U_{k,\rho} = (-1)^{\rho}(u_{k,\rho+2} - u_{k,\rho+1} - (B_{\rho+2} - B_{\rho+1})) \tag{A5}$$

$$U_s = \frac{1}{N_3}\sum_{k=1}^{3}\frac{\alpha}{3}(1+B_k)\,U_{k,2}U_{k,3} \tag{A6}$$

$$N_2 = \left[\sum_{k=1}^{3}\frac{\alpha}{3}(1+B_k)\,(U_{k,2} - U_s U_{k,3})\right]^{1/2} \tag{A7}$$

$$N_3 = \left[\sum_{k=1}^{3}\frac{\alpha}{3}(1+B_k)\,U_{k,1}\right]^{1/2} \tag{A8}$$

with ρ a cyclic parameter taking the values 1 to 3. The fixed-point equations can then be rewritten as

$$nm_n = \sum_{\mu=1}^{n\leqslant p}\sum_{\tau=1}^{2}\langle\!\langle U_{\xi^\mu,\tau}\; I((-1)^{\tau+1}\tilde{\Lambda}_3, \tilde{\Lambda}_\tau, \Xi_\tau)\rangle\!\rangle \tag{A9}$$

$$\begin{aligned} C = \sum_{k=1}^{3}\sqrt{\frac{1+B_k}{6\pi\alpha r}}\sum_{\tau=1}^{2}\Bigg\langle\!\Bigg\langle U_{k,\tau}\Bigg\{\left[\frac{T_{k,2} - (-1)^{\tau+1}\Xi_\tau\, T_{k,1}}{\sqrt{1+\Xi_\tau^2}}\right] \\ \times E\left((-1)^{\tau+1}\frac{\tilde{\Lambda}_\tau}{\sqrt{1+\Xi_\tau^2}}, (-1)^{\tau+1}\sqrt{1+\Xi_\tau^2}, \frac{\Xi_\tau\tilde{\Lambda}_\tau}{\sqrt{1+\Xi_\tau^2}}\right) \\ + (-1)^{\tau}T_{k,1}E(\tilde{\Lambda}_3, (-1)^{\tau}\Xi_1, -\tilde{\Lambda}_\tau)\Bigg\}\Bigg\rangle\!\Bigg\rangle \end{aligned} \tag{A10}$$

$$\begin{aligned} r = \frac{1}{(1-C)^2}\Bigg\{(1+B_3)(2-B_3) + \sum_{\tau=1}^{2}(-1)^{\tau+1}(1 - B_{\tau+1} - B_{\tau+2}) \\ \times (B_{\tau+1} - B_{\tau+2})\,\langle\!\langle I((-1)^{\tau+1}\tilde{\Lambda}_3, \tilde{\Lambda}_\tau, \Xi_\tau)\rangle\!\rangle\Bigg\} \end{aligned} \tag{A11}$$

where

$$\tilde{\Lambda}_\tau = (-1)^{\tau+1}\Lambda_{\tau+1,\tau+2}(\sqrt{\tau}\,\bar{\delta}_{\tau,3}N_2 + \sqrt{\tau}\,\delta_{\tau,3}N_3)^{-1} \qquad \bar{\delta} = 1-\delta \tag{A12}$$

with $\tau = 1$ to 3 cyclic and

$$\Lambda_{\rho,\sigma} = \frac{\alpha}{2}\frac{C}{1-C}(B_\rho - B_\sigma)(1-B_\rho-B_\sigma) + m_n\sum_{\mu=1}^{n}[u_{\xi^\mu,\rho} - u_{\xi^\mu,\sigma} - (B_\rho - B_\sigma)] \tag{A13}$$

$$\Xi_1 = \frac{N_3}{N_2}(1-U_s) \qquad \Xi_2 = \frac{N_3}{N_2}U_s \tag{A14}$$

$$I(x,y,z) = \frac{1}{2}\int_x^\infty \mathrm{D}t\left[1+\operatorname{erf}\left(\frac{y+zt}{\sqrt{2}}\right)\right] \tag{A15}$$

$$E(x,y,z) = \frac{1}{2}\left[1-\operatorname{erf}\left(\frac{y\tilde{\Lambda}_3+z}{\sqrt{2}}\right)\right]\mathrm{e}^{-x^2/2}. \tag{A16}$$

For zero bias, these equations are equivalent to that derived in [1].

References

[1] Kanter I 1988 *Phys. Rev.* A **37** 2739
[2] Bollé D and Mallezie F 1989 *J. Phys. A: Math. Gen.* **22** 4409
[3] Bollé D, Dupont P and Mallezie F 1989 *Neural networks and spin glasses* ed W K Theumann and R Köberle (Singapore: World Scientific) p 151
[4] Bollé D and Dupont P 1990 *Statistical Mechanics of Neural Networks (Lecture Notes in Physics 368)* ed L Garrido (Berlin: Springer) p 365
[5] Bollé D, Dupont P and van Mourik J 1991 *J. Phys. A: Math. Gen.* **24** 1065
[6] Nadal J P and Rau A 1991 *J. Physique* I **1** 1109
[7] Watkin T, Rau A, Bollé D and van Mourik J 1992 *J. Physique* I **2** 167
[8] Shim G M, Kim D and Choi Y M 1991 *Phys. Rev.* A **43** 7012
[9] Shim G M, Kim D and Choi Y M 1991 *Phys. Rev.* A **45** 1238
[10] Vogt H and Zippelius A 1992 *J. Phys. A: Math. Gen.* **25** 2209
[11] Patrick A E, Picco P, Ruiz J and Zagrebnov V A 1991 *J. Phys. A: Math. Gen.* **24** L637
[12] Bollé D, Vinck B and Zagrebnov V A 1992 On parallel dynamics of the Q-state Potts and Q-Ising neural networks *Preprint* KUL-TF 92/12
[13] Ferrari P A, Martinez S and Picco P 1992 *J. Stat. Phys.* **66** 1643
[14] Gayrard V 1991 The thermodynamic limit of the Q-state Potts–Hopfield model with infinitely many patterns *Preprint*
[15] Bollé D, Dupont P and Huyghebaert J 1992 *Phys. Rev.* A **46** 4194; 1992 *Physica A* in press
[16] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530; 1987 *Ann. Phys., NY* **173** 30
[17] Amit D J, Gutfreund H and Sompolinsky H 1987 *Phys. Rev.* A **35** 2293
[18] Doedel E 1981 *Cong. Num.* **30** 265
[19] de Almeida J R and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
[20] Cools R and Haegemans A 1992 *NATO ASI Series C* vol 357, ed T O Espelid and A Genz (Dordrecht: Kluwer) p 305
[21] Berntsen J and Espelid T O 1992 *ACM Trans. Math. Soft.* in press
[22] Naef J P and Canning A 1992 *J. Physique* I **2** 247