Thouless-Anderson-Palmer equations for neural networks

Maoz Shamir and Haim Sompolinsky
Racah Institute of Physics and Center for Neural Computation, The Hebrew University, Jerusalem 91904, Israel

(Received 28 June 1999)

Previous derivation of the Thouless-Anderson-Palmer (TAP) equations for the Hopfield model by the cavity method yielded results that were inconsistent with those of the perturbation theory as well as the results derived by the replica theory of the model. Here we present a derivation of the TAP equation for the Hopfield model by the cavity method and show that it agrees with the form derived by perturbation theory. We also use the cavity method to derive TAP equations for the pseudoinverse neural network model. These equations are consistent with the results of the replica theory of these models.

PACS number(s): 87.18.Sn, 75.10.Nr, 84.35.+i

I. INTRODUCTION

Neural network models have been studied extensively using statistical mechanical methods developed for the mean-field theory of spin glasses. Amit, Gutfreund, and Sompolinsky [1] have applied the replica method [2] for the investigation of the Hopfield model [3]. The complementary approach of Thouless, Anderson, and Palmer [4] (TAP) was applied to the Hopfield model by Mézard, Parisi, and Virasoro [5], who have used the cavity method to derive TAP equations for the model. This method consists of two steps. First, a new spin is added to the system, and the distribution of the local field induced on it is characterized, in terms of the variance of the overlaps of the system states with the memorized patterns. This variance is evaluated by adding a new pattern to the system. In Ref. [5] the cavity method was applied using certain assumptions about the ultrametric structure of the phase space of the system. However, the TAP equations derived in Ref. [5] are inconsistent with the predictions of the replica solution of the Hopfield model [1]. In particular, the two theories yield different values of the transition temperature of the model. This last problem has been noted recently by Nakanishi and Takayama [6]. They presented a derivation of TAP equations for the Hopfield model, following the method introduced by Plefka [7] for the Sherrington-Kirkpatrick (SK) spin-glass model. This method is based on an expansion of the Gibbs potential in powers of the exchange coupling. The TAP equations derived by Nakanishi and Takayama differed from those of Mézard *et al.* and are similar to those presented previously by Fukai and Shiino [8], in particular, they predict a transition temperature that agrees with the replica solution. The origin of the discrepancy between the two derivations of TAP equations remained unclear.

In this paper we reexamine the derivation of the TAP equations by the cavity method. Our goals are first, to develop an appropriate cavity method that does not depend on additional ultrametric assumptions; second, to resolve the apparent discrepancy between the cavity method and the results derived by perturbation theory as well as by the replica theory. Finally, we will use the cavity method to derive the form of the TAP equations for the more complex pseudoinverse model [9] of associative memory. This model has been investigated previously by the replica theory only. The outline of the paper is as follows. We begin by describing the cavity method for the relatively simple case of the SK infinite-range spin glass model [10]. In Sec. III we extend the method to derive TAP equations for the Hopfield model, and show that our results are in agreement with the equations derived by Nakanishi and Takayma. In Sec. IV the TAP equations for the pseudoinverse model [9] are derived. Our conclusions are presented in the last section.

II. TAP EQUATIONS FOR THE SHERRINGTON-KIRKPATRICK MODEL

A. Definition of the model

The model system is a system of N Ising spins governed by a Hamiltonian

$$H^{(N)}=-\frac{1}{2}\sum_{i,j=1}^{N} J_{ij}s_i s_j . \tag{1}$$

The upper index (N) denotes that it relates to a system with N spins. The J_{ij}'s are independent random Gaussian variables, distributed according to

$$P(J_{ij})=\frac{\sqrt{N}}{\sqrt{2\pi J}}\exp\left(-\frac{1}{2}\frac{NJ_{ij}^2}{J^2}\right) \tag{2}$$

and $J_{ij}=J_{ji}$.

B. Adding a spin to the system

Following Ref. [5] we add a spin to the system and calculate its thermal average in the $(N+1)$-spin system as a function of averages in the N-spin system. Adding a spin s_0 at site zero, we also add a set of interaction constants $\{J_{0j}\}_{j=1}^{N}$ that are distributed according to Eq. (2). The Hamiltonian of the $(N+1)$-spin system is defined

$$H^{(N+1)}=H^{(N)}-h_0 s_0 , \tag{3}$$

$$h_0=\sum_{j=1}^{N} J_{0j}s_j . \tag{4}$$

The states of the system are distributed according to a Gibbs distribution with a Hamiltonian $H^{(N+1)}$,

$$P^{N+1}(\{s_i\}_{i=0}^{N})=\frac{1}{Z_{N+1}}\exp(-\beta H^{(N+1)}), \tag{5}$$

$$Z_{N+1}=\mathrm{Tr}_{\{s_i\}_{i=0}^{N}}\ \exp(-\beta H^{(N+1)}). \tag{6}$$

From the distribution of states of the $(N+1)$-spin system, Eq. (5), we obtain the joint probability distribution of the local field and spin at site zero,

$$P^{N+1}(h_0,s_0)=\frac{1}{Z_{N+1}}\mathrm{Tr}_{\{s_i\}_{i=1}^{N}}\left[\delta\left(h_0-\sum_{j=1}^{N}J_{0j}s_j\right)\times\exp(-\beta H^{(N+1)})\right]. \tag{7}$$

The dependence on s_0 is via Eq. (3). Introducing

$$P^{N}(h_0)=\frac{1}{Z_N}\mathrm{Tr}_{\{s_i\}_{i=1}^{N}}\left[\delta\left(h_0-\sum_{j=1}^{N}J_{0j}s_j\right)\exp(-\beta H^{(N)})\right], \tag{8}$$

Eq. (7) can be written as

$$P^{N+1}(h_0,s_0)=\frac{1}{\zeta}\exp(\beta h_0 s_0)P^{N}(h_0), \tag{9}$$

$$\zeta=\frac{Z_{N+1}}{Z_N}=\langle 2\cosh\beta h_0\rangle_N. \tag{10}$$

We use $\langle\cdots\rangle_N$ to denote thermal averaging with respect to the N-spin system. Using Eq. (9), the thermal average of the spin at site zero is given by

$$\langle s_0\rangle_{N+1}=\mathrm{Tr}_{s_0}\ s_0\int P^{N+1}(h_0,s_0)dh_0=\frac{\langle\sinh\beta h_0\rangle_N}{\langle\cosh\beta h_0\rangle_N}. \tag{11}$$

Similarly,

$$\langle h_0\rangle_{N+1}=\frac{\langle h_0\cosh\beta h_0\rangle_N}{\langle\cosh\beta h_0\rangle_N}. \tag{12}$$

C. Statistics of the local field

The first two moments of the local field at site zero in the N-spin system are

$$\langle h_0\rangle_N=\sum_{j=1}^{N}J_{0j}\langle s_j\rangle_N, \tag{13}$$

$$\langle(\delta h_0)^2\rangle_N=\sum_{i,j=1}^{N}J_{0i}J_{0j}\langle\delta s_i\delta s_j\rangle_N, \tag{14}$$

where $\delta s_i\equiv s_i-\langle s_i\rangle$. The J_{0j}'s are random independent variables of the order of $1/\sqrt{N}$ with zero mean. For $i\neq j$, $\langle\delta s_i\delta s_j\rangle_N$ is of the order of $1/\sqrt{N}$. Since $\langle\delta s_i\delta s_j\rangle_N$ and J_{0j} are independent, the contribution of the $i\neq j$ terms in Eq. (14) is of the order of $1/\sqrt{N}$. We can, therefore, approximate Eq. (14) by the $i=j$ terms

$$\langle(\delta h_0)^2\rangle_N=\sum_{i=1}^{N}J_{0i}^2\langle(\delta s_i)^2\rangle_N=J^2(1-q_N), \tag{15}$$

$$q_N=\frac{1}{N}\sum_{i=1}^{N}\langle s_i\rangle_N^2. \tag{16}$$

The last equality in Eq. (15) results from self-averaging in the large-N limit. We now assume that in the N-spin system the local field at site zero is a Gaussian random variable. This assumption is supported by the fact that in the N-spin system h_0 is a sum of N independent random variables. We further assume that for large N we can replace q_N by its value in the thermodynamic limit, i.e., that $q_N=q$. We can, therefore, write

$$P^{N}(h_0)=\frac{1}{\sqrt{2\pi J^2(1-q)}}\exp\left(-\frac{(h_0-\langle h_0\rangle_N)^2}{2J^2(1-q)}\right). \tag{17}$$

D. TAP equations for the local magnetization

Substituting Eq. (17) into Eqs. (11) and (12), we obtain

$$\langle s_0\rangle_{N+1}=\tanh[\beta\langle h_0\rangle_N], \tag{18}$$

$$\langle h_0\rangle_{N+1}=\langle h_0\rangle_N+\beta J^2(1-q)\langle s_0\rangle_{N+1}. \tag{19}$$

Substituting Eq. (19) into Eq. (18), we retrieve the known TAP equations for the SK model ([4,5])

$$\langle s_i\rangle=\tanh\left[\beta\left(\sum_{j\neq i}J_{ij}\langle s_j\rangle-\beta J^2(1-q)\langle s_i\rangle\right)\right]. \tag{20}$$

III. TAP EQUATIONS FOR THE HOPFIELD MODEL

A. Definition of the model

The model system is a system of N binary neurons that stores p memory patterns $\{\xi_i^\mu\}$ $(i=1,\ldots,N,\mu=1,\ldots,p)$ in the connection matrix. The Hamiltonian of the system is

$$H^{(N)}=-\frac{1}{2}\sum_{i,j=1}^{N}J_{ij}s_is_j, \tag{21}$$

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu. \tag{22}$$

The ξ's are independent random binary variables $\xi_i^\mu=\pm 1$ with zero mean. We are interested in the limit of $N\rightarrow\infty$ and $p\rightarrow\infty$, such that the ratio $\alpha=p/N$ remains finite.

B. Adding a neuron to the system

The first step of the derivation of TAP equations for this model is to add a neuron s_0 at site zero and to add $\{\xi_0^\mu\}_{\mu=1}^{p}$ to the p patterns. The Hamiltonian of the $(N+1)$-neuron system is

$$H^{(N+1)}=H^{(N)}-h_0s_0, \tag{23}$$

$$h_0=\sum_{j=1}^{N}J_{0j}s_j, \tag{24}$$

$$J_{0j}=\frac{1}{N}\sum_{\mu=1}^{p}\xi_0^{\mu}\xi_j^{\mu}. \tag{25}$$

As in the SK model the joint probability of the local field and the neuron state at site zero can be written as

$$P^{N+1}(h_0,s_0)=\frac{1}{\zeta}\exp(\beta h_0 s_0)P^{N}(h_0), \tag{26}$$

$$\zeta=\frac{Z_{N+1}}{Z_N}=\langle 2\cosh\beta h_0\rangle_N. \tag{27}$$

Hence, Eqs. (11) and (12) hold also for this model.

C. Statistics of the local field

The mean and variance of the local field are

$$\langle h_0\rangle_N=\sum_{j=1}^{N}J_{0j}\langle s_j\rangle_N, \tag{28}$$

$$\langle(\delta h_0)^2\rangle_N=\sum_{i,j=1}^{N}J_{0i}J_{0j}\langle\delta s_i\delta s_j\rangle_N=\sum_{\mu,\nu}\xi_0^{\mu}\xi_0^{\nu}\langle\delta m_{\mu}\delta m_{\nu}\rangle_N, \tag{29}$$

where m_{μ} is the overlap with pattern $\{\xi_i^{\mu}\}_{i=1}^{N}$,

$$m_{\mu}=\frac{1}{N}\sum_{i=1}^{N}\xi_i^{\mu}s_i. \tag{30}$$

For $\mu\neq\nu$ we have $\langle\delta m_{\mu}\delta m_{\nu}\rangle=O(1/N^{3/2})$, hence we can approximate Eq. (29) by the contributions of the $\mu=\nu$ terms, i.e.,

$$\langle(\delta h_0)^2\rangle_N=\sum_{\mu}\langle(\delta m_{\mu})^2\rangle_N\equiv r_N. \tag{31}$$

Since $\{\xi_0^{\mu}\}$ are random and independent of the distribution of states in the N-neuron system we can approximate P^N by

$$P^{N}(h_0)=\frac{1}{\sqrt{2\pi r}}\exp\left(-\frac{(h_0-\langle h_0\rangle_N)^2}{2r}\right), \tag{32}$$

where r is the large-N limit of r_N. Substituting Eq. (32) into Eqs. (11) and (12), we obtain

$$\langle s_0\rangle_{N+1}=\tanh[\beta\langle h_0\rangle_N], \tag{33}$$

$$\langle h_0\rangle_{N+1}=\langle h_0\rangle_N+\beta r\langle s_0\rangle_{N+1}. \tag{34}$$

Substituting Eq. (34) into Eq. (33), we obtain the TAP equations for local magnetization of the Hopfield model

$$\langle s_i\rangle=\tanh\left[\beta\left(\sum_{j\neq i}J_{ij}\langle s_j\rangle-\beta r\langle s_i\rangle\right)\right]. \tag{35}$$

D. Adding a memory pattern to the Hamiltonian

In order to evaluate r we use the cavity method a second time, this time by adding a memory pattern to the Hamiltonian [5]. We define H_p to be the Hamiltonian of a system with N neurons and p memory patterns

$$H_p=-\frac{1}{2}\sum_{i,j}\left(\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^{\mu}\xi_j^{\mu}\right)s_is_j. \tag{36}$$

Adding pattern $\{\xi_i^0\}_{i=1}^{N}$ to the Hamiltonian, we define

$$H_{p+1}=H_p-\frac{1}{2N}\sum_{i,j}\xi_i^0\xi_j^0 s_is_j=H_p-\frac{1}{2}N(m_0)^2, \tag{37}$$

where m_0 is the overlap of the state with the new pattern. The probability distribution of m_0, with respect to thermal fluctuations in the system governed by H_{p+1}, can be written in the form of

$$P^{p+1}(m_0)=\frac{1}{\zeta'}\exp\left(\frac{1}{2}\beta N(m_0)^2\right)P^{p}(m_0). \tag{38}$$

In the system with H_p we have

$$\langle m_0\rangle_p=0, \tag{39}$$

$$\langle(\delta m_0)^2\rangle_p=\frac{1}{N^2}\sum_{i,j}\xi_i^0\xi_j^0\langle\delta s_i\delta s_j\rangle_p=\frac{1}{N}(1-q). \tag{40}$$

Assuming $P^p(m_0)$ is Gaussian, we obtain

$$P^{p+1}(m_0)=\frac{1}{\zeta'}\exp\left[-\frac{1}{2}N\left(\frac{1}{1-q}-\beta\right)m_0^2\right] \tag{41}$$

yielding

$$\langle(\delta m_0)^2\rangle_{p+1}=\frac{1}{N}\frac{1-q}{1-\beta(1-q)}. \tag{42}$$

Substituting Eq. (42) for each of the terms in Eq. (31) yields

$$r=\frac{\alpha(1-q)}{1-\beta(1-q)}. \tag{43}$$

Equations (35) and (43) agree with the result of Fukai and Shiino [8] and Nakanishi and Takayama [6].

IV. TAP EQUATIONS FOR THE PSEUDOINVERSE MODEL

A. Definition of the model

As in the Hopfield model, the interaction matrix is designed to store p-binary memory patterns $\{\xi_i^{\mu}\}$ $i=1,\ldots,N$, $\mu=1,\ldots,p$. The ξ's are independent random binary variables with zero mean. We start by defining the N-neuron system

$$H^{(N)}=-\frac{1}{2}\sum_{i,j=1}^{N}J_{ij}^{(N)}s_is_j, \tag{44}$$

$$J_{ij}^{(N)}=\frac{1}{N}\sum_{\mu,\nu}\xi_i^{\mu}[C^{(N)}]_{\mu\nu}^{-1}\xi_j^{\nu}, \tag{45}$$

$$C_{\mu\nu}^{(N)}=\frac{1}{N}\sum_{i=1}^{N}\xi_i^{\mu}\xi_i^{\nu}. \tag{46}$$

As shown in Ref. [9] the Hamiltonian can be written as

$$H^{(N)}=-\frac{N}{2}\sum_{\mu}m_{\mu}a_{\mu}, \tag{47}$$

where

$$a_{\mu}=\sum_{\nu}[C^{(N)}]_{\mu\nu}^{-1}m_{\nu}. \tag{48}$$

B. Adding a neuron to the system

Adding a neuron at site zero and $\{\xi_0^{\mu}\}_{\mu=1}^{p}$ we define

$$H^{(N+1)}=-\frac{1}{2}\sum_{i,j=0}^{N}J_{ij}^{(N+1)}s_is_j, \tag{49}$$

$$J_{ij}^{(N+1)}=\frac{1}{N+1}\sum_{\mu,\nu}\xi_i^{\mu}[C^{(N+1)}]_{\mu\nu}^{-1}\xi_j^{\nu}, \tag{50}$$

$$C_{\mu\nu}^{(N+1)}=\frac{1}{N+1}\sum_{i=0}^{N}\xi_i^{\mu}\xi_i^{\nu}. \tag{51}$$

We observe that

$$[C^{(N+1)}]^{-1}=\frac{N+1}{N}\Bigg([C^{(N)}]^{-1}-\frac{1}{1+\gamma}\frac{1}{N}\times[C^{(N)}]^{-1}\vec{\xi}_0\vec{\xi}_0^{t}[C^{(N)}]^{-1}\Bigg), \tag{52}$$

$$\gamma=\frac{1}{N}\vec{\xi}_0^{t}[C^{(N)}]^{-1}\vec{\xi}_0, \tag{53}$$

where $\vec{\xi}_0$ is a p-dimensional column vector of the memory patterns at site zero, and $\vec{\xi}_0^{t}$ is its transpose row vector. In Appendix A we show that

$$\gamma=\frac{\alpha}{1-\alpha}, \tag{54}$$

where $\alpha=p/N$. We denote

$$h_0^{(N)}=\frac{1}{N}\sum_{j=1}^{N}\sum_{\mu\nu}\xi_0^{\mu}[C^{(N)}]_{\mu\nu}^{-1}\xi_j^{\nu}s_j=\sum_{\mu}\xi_0^{\mu}a_{\mu}, \tag{55}$$

where the upper index (N) here indicates the use of $C^{(N)}$ and not $C^{(N+1)}$. Using Eq. (52), $H^{(N+1)}$ can be written as

$$H^{(N+1)}=H^{(N)}+\frac{(h_0^{(N)})^2}{2(1+\gamma)}-\frac{h_0^{(N)}s_0}{1+\gamma}, \tag{56}$$

hence

$$P^{N+1}(h_0^{(N)},s_0)=\frac{1}{\zeta}\exp\left(-\frac{\beta(h_0^{(N)})^2}{2(1+\gamma)}+\frac{\beta h_0^{(N)}s_0}{1+\gamma}\right)P^{N}(h_0^{(N)}), \tag{57}$$

where ζ is a normalization constant. Using Eq. (57) we obtain

$$\langle s_0\rangle_{N+1}=\frac{1}{\xi}\left\langle 2\exp\left(-\frac{\beta(h_0^{(N)})^2}{2(1+\gamma)}\right)\sinh\left(\frac{\beta h_0^{(N)}s_0}{1+\gamma}\right)\right\rangle_N, \tag{58}$$

$$\langle h_0^{(N)}\rangle_{N+1}=\frac{1}{\xi}\left\langle 2h_0^{(N)}\exp\left(-\frac{\beta(h_0^{(N)})^2}{2(1+\gamma)}\right)\cosh\left(\frac{\beta h_0^{(N)}}{1+\gamma}\right)\right\rangle. \tag{59}$$

C. Statistics of the local field

The first two moments of $h_0^{(N)}$ are

$$\langle h_0^{(N)}\rangle_N=\sum_{\mu}\xi_0^{\mu}\langle a_{\mu}\rangle_N, \tag{60}$$

$$\langle(\delta h_0^{(N)})^2\rangle_N=\sum_{\mu,\nu}\xi_0^{\mu}\xi_0^{\nu}\langle\delta a_{\mu}\delta a_{\mu}\rangle_N=\sum_{\mu}\langle(\delta a_{\mu})^2\rangle_N\equiv Tx_N. \tag{61}$$

Assuming that $P^N(h_0^{(N)})$ is Gaussian,

$$P^N(h_0^{(N)})=\frac{1}{\sqrt{2\pi Tx}}\exp\left(-\frac{(h_0^{(N)}-\langle h_0^{(N)}\rangle_N)^2}{2Tx}\right), \tag{62}$$

where x is the large-N limit of x_N. We can calculate $\langle s_0\rangle_{N+1}$ and $\langle h_0^{(N)}\rangle_{N+1}$ using Eqs. (58) and (59),

$$\langle s_0\rangle_{N+1}=\tanh\left(\frac{\beta\langle h_0^{(N)}\rangle_N}{1+\gamma+x}\right), \tag{63}$$

$$\langle h_0^{(N)}\rangle_{N+1}=\frac{1+\gamma}{1+\gamma+x}\langle h_0^{(N)}\rangle_N+\frac{x}{1+\gamma+x}\langle s_0\rangle_{N+1}. \tag{64}$$

D. TAP equations for the pseudoinverse model

Using Eq. (52) we obtain an expression for the local field at site zero in the $(N+1)$-neuron system

$$\langle h_0^{(N+1)}\rangle_{N+1}=\frac{1}{N+1}\sum_{j=1}^{N}\sum_{\mu,\nu}\xi_0^{\mu}[C^{(N+1)}]_{\mu\nu}^{-1}\xi_j^{\nu}\langle s_j\rangle_{N+1}=\frac{1}{1+\gamma}\langle h_0^{(N)}\rangle_{N+1}. \tag{65}$$

Using Eqs. (64) and (65), we obtain

$$\langle h_0^{(N+1)}\rangle_{N+1}=\frac{\langle h_0^{(N)}\rangle_N}{1+\gamma+x}+\frac{\langle s_0\rangle_{N+1}}{1+\gamma+\dfrac{(1+\gamma)^2}{x}}. \tag{66}$$

Substituting Eq. (66) into Eq. (63), we obtain the TAP equations for the pseudoinverse model

$$\langle s_i\rangle = \tanh\left[\beta\left(\sum_{j\neq i} J_{ij}\langle s_j\rangle - \frac{x}{(1+\gamma)(1+\gamma+x)}\langle s_i\rangle\right)\right]. \tag{67}$$

E. Adding a memory pattern to the Hamiltonian

The evaluation of x is done by using the cavity method a second time, adding a memory pattern to the Hamiltonian. Details of the calculation are explained in Appendix B. The result of the calculation yields

$$x = \frac{C-1+\sqrt{(1-C)^2+4\alpha C}}{C+1-\sqrt{(1-C)^2+4\alpha C}} - \gamma, \tag{68}$$

where

$$C = \beta(1-q). \tag{69}$$

V. DISCUSSION

Previous application of the cavity method to the Hopfield model [5] yielded TAP equations with a cavity term that was in disagreement with the equations derived by perturbation theory [6]. Mézard *et al.* [5] applied the cavity method on soft variables generated by a Hubbard-Stantonovitch transformation of the original Ising system. However, the relation between the statistics of the soft variables and the spin variables must be treated with care. If these relations are taken appropriately, then their method yields the same equations as derived by perturbation theory. Here we have avoided using the Hubbard-Stantonovitch transformation all together and applied the cavity method directly on the Ising spin system. In addition, we have shown that the correct TAP equations can be derived by the cavity method without additional assumptions about the structure of the minima or their energy distribution.

We now briefly discuss the correspondence between the replica theory and the TAP equations. In the SK model, this correspondence has been extensively studied [5]. Assuming an ultrametric structure of the TAP solutions yields a mean field theory that is equivalent to Parisi's replica solution [5]. A similar study for the neural network models has not been made. Here we note two points of agreement between the theories. Equations (35) and (43) for the Hopfield model predict a second-order transition from a paramagnetic state $\langle s_i\rangle = 0$ to a spin-glass state in which $\langle s_i\rangle$ are different from zero but they do not have a macroscopic overlap with any of the patterns. This transition occurs at a temperature $T_g = 1+\sqrt{\alpha}$, as was shown by Nakanishi and Takayama [6], which agrees with the replica theory [1]. In the case of the pseudo-inverse model, Eqs. (67), (45), and (46) admit a solution of the form

$$\langle s_i\rangle = m\xi_i^{\mu}. \tag{70}$$

This corresponds to a retrieval state in which the sign of the local magnetizations is identical to the pattern μ. Substituting this ansatz in Eq. (67), and using Eqs. (B15) and (A4) yields the following mean-field equation for m:

$$m = \tanh(\beta J m), \tag{71}$$

where

$$J = \frac{1}{2C}\left[1+C-\sqrt{(1-C)^2+4\alpha C}\right]. \tag{72}$$

These equations agree with the results of the replica theory for the retrieval state in this model, see Eqs. (3.12) and (3.8) in Ref. [9]. In conclusion, we believe that the TAP equations derived here for neural network models of associative memory are equivalent to the replica theory for these models.

ACKNOWLEDGMENTS

This research is partially supported by a grant from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.

APPENDIX A: ESTIMATION OF γ

We now calculate the value of γ for a typical choice of ξ's. For large N, we can evaluate

$$\gamma = \frac{1}{N}\,\mathrm{Tr}[C^{-1}] \tag{A1}$$

To calculate the trace of C^{-1}, we use the result for the eigenvalues spectrum of the Hopfield matrix; for $J_{ij} = (1/N)\Sigma_{\mu=1}^{\alpha N}\xi_i^{\mu}\xi_j^{\mu}(1-\delta_{ij})$ we have

$$\rho^J(\lambda) = \begin{cases} \rho_0^J(\lambda) + (1-\alpha)\delta(\lambda+\alpha), & \alpha \leqslant 1 \\ \rho_0^J(\lambda), & \alpha > 1, \end{cases} \tag{A2}$$

$$\rho_0^J(\lambda) = \frac{\sqrt{4\alpha-(1-\lambda)^2}}{2\pi(\lambda+\alpha)}. \tag{A3}$$

In our case, we take $\alpha' = 1/\alpha$ and add the diagonal term. Performing some algebra we obtain

$$\gamma = \frac{\alpha}{1-\alpha}. \tag{A4}$$

APPENDIX B: CALCULATION OF X

The partition function of a system of N neurons with p-stored memory patterns takes the following form:

$$\begin{aligned} Z_p &= \mathrm{Tr}_{\{s_i\}}\exp\left(\frac{1}{2}\beta N \sum_{\mu,\nu=1}^{p} m_\mu C_{\mu\nu}^{-1} m_\nu\right) \\ &= \mathrm{Tr}_{\{s_i\}}\int \prod_{\mu=1}^{p} dx_\mu \, \exp\left(-\frac{1}{2}\sum_{\mu,\nu=1}^{p} x_\mu C_{\mu\nu} x_\nu \right. \\ &\quad \left. + \sqrt{\beta N}\sum_{\mu=1}^{p} x_\mu m_\mu\right). \end{aligned} \tag{B1}$$

Adding a new pattern $\{\xi_i^0\}_{i=1}^N$ to the Hamiltonian, we can write

$$P^{p+1}(x_0,y_0,m_0)=\frac{1}{\zeta}\exp\left(-\frac{1}{2}x_0^2-x_0y_0+\sqrt{\beta N}x_0m_0\right)P^p(y_0,m_0), \tag{B2}$$

where $y_0=\Sigma_{\mu=1}^{p}C_{0\mu}x_\mu$. Assuming $P^p(y_0,m_0)$ is Gaussian we can write

$$P^p(\delta y_0,m_0)=\frac{1}{\zeta'}\exp\left[-\frac{1}{2}(\delta y_0,m_0)\begin{pmatrix}a & b\\ b & c\end{pmatrix}\begin{pmatrix}\delta y_0\\ m_0\end{pmatrix}\right], \tag{B3}$$

$$\begin{pmatrix}a & b\\ b & c\end{pmatrix}^{-1}=\begin{pmatrix}\langle(\delta y_0)^2\rangle_p & \langle\delta y_0\delta m_0\rangle_p\\ \langle\delta y_0\delta m_0\rangle_p & \langle(\delta m_0)^2\rangle_p\end{pmatrix}, \tag{B4}$$

where $\delta y_0\equiv y_0-\langle y_0\rangle_p$. We expect m_0 not to condense, i.e., $\langle m_0\rangle_p=0$. From Eq. (B1), $\{x_\mu\}$ are random variables with Gaussian fluctuations around a_μ, hence

$$\langle\delta x_\mu\delta x_\nu\rangle_p=\beta N\langle\delta a_\mu\delta a_\nu\rangle_p+C_{\mu\nu}^{-1}, \tag{B5}$$

$$\langle\delta x_\mu\delta m_\mu\rangle_p=\sqrt{\beta N}\langle\delta a_\mu\delta m_\mu\rangle_p. \tag{B6}$$

Using Eqs. (B5) and (B6) we obtain

$$\langle(\delta m_0)^2\rangle_p=\frac{1}{N^2}\sum_{i,j}\xi_i^0\xi_j^0\langle\delta s_i\delta s_j\rangle_p=\frac{1}{N}(1-q), \tag{B7}$$

$$\begin{aligned}\langle(\delta y_0)^2\rangle_p&=\sum_{\mu,\nu}\frac{1}{N^2}\sum_{i,j}\xi_i^0\xi_j^0\xi_i^\mu\xi_j^\nu\langle\delta x_\mu\delta x_\nu\rangle_p\\ &=\frac{1}{N}\sum_{\mu,\nu}C_{\mu\nu}\langle\delta x_\mu\delta x_\nu\rangle_p\\ &=\beta\sum_\mu\langle\delta m_\mu\delta a_\mu\rangle_p+\alpha,\end{aligned} \tag{B8}$$

$$\begin{aligned}\langle\delta y_0\delta m_0\rangle_p&=\frac{1}{N^2}\sum_{i,j,\mu}\xi_i^0\xi_j^0\xi_j^\mu\langle\delta x_\mu\delta s_i\rangle_p\\ &=\sum_\mu\frac{1}{N}\langle\delta x_\mu\delta m_\mu\rangle_p=\sqrt{\beta/N}\sum_\mu\langle\delta m_\mu\delta a_\mu\rangle_p.\end{aligned} \tag{B9}$$

Integrating over y_0 in Eq. (B2) yields

$$P^{p+1}(\delta x_0,\delta m_0)=\frac{1}{\zeta''}\exp\left[-\frac{1}{2}(\delta x_0,\delta m_0)\times\begin{pmatrix}1-\frac{1}{a} & -\sqrt{\beta N}-\frac{b}{a}\\ -\sqrt{\beta N}-\frac{b}{a} & c-\frac{b^2}{a}\end{pmatrix}\times\begin{pmatrix}\delta x_0\\ \delta m_0\end{pmatrix}\right]. \tag{B10}$$

Here $\delta m_0=m_0-\langle m_0\rangle_{p+1}$ and $\delta x_0=x_0-\langle x_0\rangle_{p+1}$. We denote $U=\beta\Sigma_\mu\langle\delta m_\mu\delta a_\mu\rangle_p$.

Using Eqs. (B7), (B8), and (B9) we calculate $\langle\delta x_0\delta m_0\rangle_{p+1}$ from Eq. (B10) in terms of U, and then solve the self-consistent equation

$$U=\alpha\sqrt{\beta N}\langle\delta x_0\delta m_0\rangle_{p+1} \tag{B11}$$

yielding

$$U=\frac{1}{2}[C-1+\sqrt{(1-C)^2+4\alpha C}], \tag{B12}$$

where

$$C=\beta(1-q). \tag{B13}$$

From Eq. (B10) we obtain

$$\langle(\delta x_0)^2\rangle=\frac{U}{\alpha(C-U)}, \tag{B14}$$

hence

$$x=\frac{1}{N}\sum_\mu\{\langle(\delta x_\mu)^2\rangle-[C^{-1}]_{\mu\mu}\}=\frac{U}{(C-U)}-\frac{1}{N}\mathrm{Tr}[C^{-1}]. \tag{B15}$$

[1] D.J. Amit, H. Geutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985); Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. (N.Y.) **17**, 22 (1987).

[2] S.F. Edwards and P.W. Anderson, J. Phys. F: Met. Phys. **5**, 965 (1975).

[3] J.J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[4] D.J. Thouless, P.W. Anderson, and R.G. Palmer, Philos. Mag. **35**, 593 (1977).

[5] M. Mézard, G. Parisi, and M.A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).

[6] K. Nakanishi and H. Takayama, J. Phys. A **30**, 8085 (1997).

[7] T. Plefka, J. Phys. A **15**, 1971 (1982).

[8] T. Fukai and M. Shiino, J. Phys. A **25**, 2873 (1992).

[9] I. Kanter and H. Sompolinsky, Phys. Rev. A **35**, 380 (1987).

[10] D. Sherrington and S. Kirkpatrick, Phys. Rev. B **17**, 4384 (1978).