Space of interactions with definite symmetry in neural networks with biased patterns as a spin-glass problem

Alba Theumann
Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixa Postal 15051, 91501-970 Porto Alegre, Rio Grande do Sul, Brazil

(Received 17 July 1995; revised manuscript received 10 January 1996)

We study the space of interactions of a connected neural network with biased patterns, when the synaptic interactions satisfy a symmetry constraint. We show that the solution to the problem requires the calculation of a quantity $N\Omega_\mu$ analogous to the thermodynamic potential of a multiply connected Ising model with site dependent interactions, which maps the present problem into the spin-glass problem. By using a diagrammatic expansion, we express $N\Omega_\mu$ formally as a functional of renormalized site dependent ''propagators'' G_{ij} and local ''magnetizations'' m_i, which are determined from a variational principle. Calculating $N\Omega_\mu$ in the single site or Brout approximation we recover the theory of Thouless, Anderson, and Palmer (TAP), while the m_i satisfy TAP-like equations. In the impossibility of solving the equations, we analyze an approximate solution that sums only tree diagrams and interpolates between the two known results of total asymmetry, finite bias, and arbitrary symmetry with vanishing bias. The results show a small dependence on the asymmetry parameter. [S1063-651X(96)11006-0]

PACS number(s): 87.10.+e, 64.10.+h, 64.60.Cn, 05.50.+q

I. INTRODUCTION

In her classical seminal work [1], Gardner proposed and solved analytically the problem of calculating the fractional volume of the space of interactions that allows a set of patterns $\{\xi_i^\mu\}$ to be fixed points of the dynamical equations, for a network of neurons $S_i=\pm 1$ located at sites $i=1\ldots N$,

$$S_i(t+1)=\mathrm{sgn}\left[\frac{1}{\sqrt{N}}\sum_j{}' J_{ij}S_j(t)-T_i\right], \tag{1}$$

where the primed sum indicates the restriction to $j\neq i$. In Eq. (1) the asymmetric interactions $J_{ij}\neq J_{ji}$ satisfy the spherical condition

$$\frac{1}{N}\sum_j{}' J_j^2=1 \tag{2}$$

and T_i is an arbitrary threshold. The problem then consists of calculating the volume of the space of interactions J_{ij} such that the set of N equations

$$\xi_i^\mu=\mathrm{sgn}\left[\frac{1}{\sqrt{N}}\sum_j{}' J_{ij}\xi_j^\mu-T_i\right], \quad i=1\ldots N \tag{3}$$

are simultaneously satisfied for every pattern $\{\xi_i^\mu\}$, where $\xi_i^\mu=\pm 1$ and $\mu=1,\ldots,p$. We call $b=\langle\xi_i^\mu\rangle$ the pattern's bias.

The elegance and simplicity of Gardner's method stems from the fact that when J_{ij} and J_{ji} are independent variables the N equations (3) also decouple and the volume in phase space for each set of interactions J_{ij} starting with a given value of i can be calculated exactly either for biased or unbiased patterns [1].

In a subsequent work [2] a more general problem was treated, namely that the interactions J_{ij} and J_{ji} are no longer independent but satisfy a symmetry constraint. This is a far more complicated case because the N equations (3) no longer decouple and in Ref. [2] the problem was solved only for unbiased patterns in a strongly diluted lattice with connectivity $C\leqslant\ln(N)$, within a replica symmetric and site independent theory.

In the present paper we study the space of interactions of a connected network with arbitrary values of $C\leqslant N$ and with biased patterns, when the interactions J_{ij} and J_{ji} are linked by a symmetry constraint. We show that the solution to the problem requires the calculation of a quantity $N\Omega_\mu$ analogous to the thermodynamic potential of a multiply connected Ising model with site dependent interactions, which maps the present problem into the long range spin-glass problem [3] that was studied by means of diagrammatic methods by Sommers [4].

We follow the diagrammatic techniques of Ref. [4] together with the inclusion of Lagrange multipliers that insure the conservation of the local identity $\xi_i^2=1$, generalizing the derivation of Thouless, Anderson, and Palmer (TAP) equations [5] by Southern and Young [6]. The linked cluster theorem allows us to express formally $N\Omega_\mu$ as a functional of renormalized site dependent ''propagators'' G_{ij} and local ''magnetizations'' m_i, which are determined from a variational principle. We show that within a single site approximation $N\Omega_\mu$ is given by TAP-like free energy while the m_i satisfy TAP-like equations [5].

In the case of *unbiased* patterns these equations accept the trivial, explicitly site independent, solution $m_i=0$ and our results coincide with those obtained in Ref. [2]. Given the impossibility of solving the equations for finite bias $b>0$, we analyze an approximate solution that interpolates between the two results in Refs. [1] and [2] when $\eta=0, b\neq 0$ and $\eta\neq 0, b=0$, respectively.

The paper is organized as follows. We describe the model in Sec. II, while we present in Sec. III a rigorous diagram-

matic analysis to derive TAP-like equations. In Sec. IV we present the results for the approximate interpolating local field, and we leave Sec. V for discussions.

II. DESCRIPTION OF THE MODEL

The problem we want to solve is the calculation of the volume in phase space occupied by interactions J_{ij} that satisfy the dynamical equations:

$$\xi_i^\mu \sum_{j\neq i} C_{ij} J_{ij} \xi_j^\mu > \kappa \tag{4}$$

where the bond occupation variable $C_{ij}=C_{ji}$, with probability distributions

$$P(C_{ij}) = \frac{C}{N}\delta(C_{ij}-1) + \left(1-\frac{C}{N}\right)\delta(C_{ij}), \tag{5}$$

and without loss of generality [1] we set $T_i=0$.

The interactions J_{ij} in Eq. (4) are not symmetric, $J_{ij} \neq J_{ji}$, and are subject to the spherical constraint

$$\sum_{j\neq i} C_{ij} J_{ij}^2 = \frac{C}{N}\sum_{j\neq i} J_{ij}^2 = C \tag{6}$$

and to the symmetry constraint

$$\sum_{j\neq i} C_{ij} J_{ij} J_{ji} = \frac{C}{N}\sum_{j\neq i} J_{ij} J_{ji} = \eta C. \tag{7}$$

For $\eta=0$, there is no correlation between J_{ij} and J_{ji}, while for $\eta=1$ we recover the symmetric situation $J_{ij}=J_{ji}$.

The patterns $\{\xi_i^\mu\}, \mu=1\ldots p$ in Eq. (4) are independent random variables with the probability distribution

$$P_0(\xi_j^\mu) = \frac{e^{\tau\xi_j^\mu}}{2\cosh(\tau)}[\delta(\xi_j^\mu-1)+\delta(\xi_j^\mu+1)], \tag{8}$$

then $(\xi_j^\mu)^2=1$ and

$$\langle \xi_j^\mu \rangle_0 = \int d\xi_j^\mu P_0(\xi_j^\mu)\xi_j^\mu = \tanh(\tau) = b, \tag{9}$$

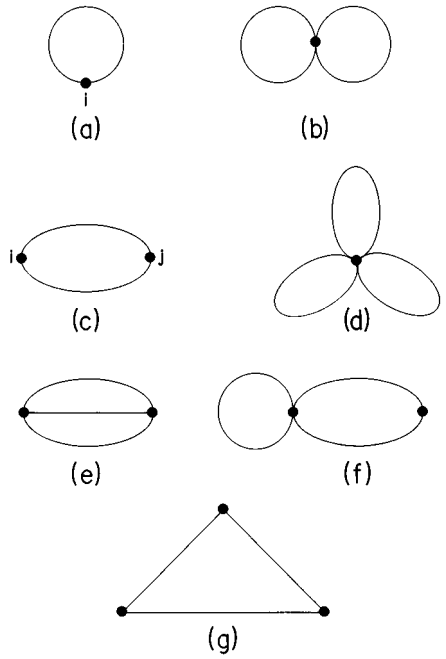

FIG. 1. Diagrammatic expansion of $N\Omega'_\mu$. A single line joining sites i and j stands for G_{ij}^0 and a dot linking r legs at site j stands for a generalized cumulant average in Eq. (A7).

where b is the bias.

In the following we will omit, for simplicity, the bond variable C_{ij}, and it is understood that a sum over the J_{ij}'s implies also a bond average, according to Eqs. (6) and (7). The volume V in the space of the J_{ij} of solutions of Eq. (1) subject to the constraints in Eqs. (6) and (7) is [1]

$$V = \prod_i \left\{ \int \prod_{j\neq i} dJ_{ij} \prod_\mu \theta\left(\xi_i^\mu \sum_{j\neq i} \frac{J_{ij}}{\sqrt{C}}\xi_j^\mu - \kappa\right) \times \delta\left(\sum_{j\neq i} J_{ij}^2 - C\right)\delta\left(\sum_{j\neq i} J_{ij}J_{ji} - \eta C\right)\right\}. \tag{10}$$

To average over the ξ_i^μ we use the replica method and we obtain [1], by using the integral representation of the Heaviside function,

$$V_n = \langle V^n \rangle = \int_\kappa^\infty \prod_{\alpha,\mu,j} d\lambda_{\mu j}^\alpha \int_{-\infty}^\infty \prod_{\alpha,\mu,j} dx_{\mu j}^\alpha \exp\left(i\sum_{\alpha,\mu,j} x_{\mu j}^\alpha \lambda_{\mu j}^\alpha\right) \int \prod_\alpha \prod_{ij} dJ_{ij}^\alpha \exp\left(\sum_\mu \Omega_\mu N\right) \times \prod_{i,\alpha}\left[\delta\left(\sum_j{}' J_{ij}^{\alpha^2} - C\right)\delta\left(\sum_j{}' J_{ij}^\alpha J_{ji}^\alpha - \eta C\right)\right] \tag{11}$$

where the replica index $\alpha=1\ldots n$ and

$$e^{\Omega_\mu N} = \left\langle \exp\left\{-i\frac{1}{2\sqrt{C}}\sum_{\langle i,j\rangle} t_{ij}^\mu \xi_i^\mu \xi_j^\mu\right\}\right\rangle_0. \tag{12}$$

The sum in Eq. (12) is over all pairs $\langle i,j\rangle, i\neq j$, and

$$t_{ij}^\mu = \sum_\alpha (x_{\mu i}^\alpha J_{ij}^\alpha + x_{\mu j}^\alpha J_{ji}^\alpha). \tag{13}$$

The bracket in Eq. (12) indicates an average over the independent variables ξ_i^μ at each site that take values ± 1 with the probability distribution in Eq. (8).

III. FORMAL EXPRESSION FOR Ω_μ

From Eqs. (8) and (12) we obtain

$$e^{N\Omega\mu}=\exp\left(-\sum_i \nu_i\right)\int\prod_j [P_0(\xi_j)d\xi_j]$$
$$\times\exp\left(-\frac{1}{2}\sum_{i,j}\xi_i\xi_j G^0_{ij}\right), \tag{14}$$

with

$$G^0_{ij}=\frac{i}{\sqrt{C}}t^\mu_{ij}-2\nu_i\delta_{ij}, \quad t^\mu_{ii}=0 \tag{15}$$

and where we introduced the identity $\Sigma_i\nu_i[\xi_i^2-1]=0$ into the exponential. The ν_i are Lagrange multipliers that will insure that the identity is preserved on the average procedure and will be chosen to minimize $N\Omega_\mu$. This is a generalization of the method used by Southern and Young [6] to derive TAP equations. From now on the prefix ''μ'' is not written explicitly unless it gives rise to confusion.

Introducing Eq. (8) into Eq. (14), we recognize that $N\Omega_\mu$ corresponds to the free energy of a multiply connected Ising model with site dependent interactions $\beta J_{ij}=-(i/\sqrt{C})t^\mu_{ij}$ in the presence of a ''magnetic field'' τ, and this problem has been analyzed diagrammatically by Sommers [4] in the study of the Sherrington-Kirkpatrick [3] model for a spin glass. We follow here a related procedure more appropriate to our problem that we describe in some detail for unfamiliar readers. We start by shifting variables in Eq. (14),

$$\xi_i=\eta_i+m_i, \tag{16}$$

where the m_i will be determined self-consistently below and we obtain

$$N\Omega_\mu=-\sum_i \nu_i-\frac{1}{2}\sum_{i,j} G^0_{ij}m_i m_j-\sum_j(\tau_j-\tau)m_j$$
$$+\sum_j \log\frac{\cosh(\tau_j)}{\cosh(\tau)}+N\Omega'_\mu, \tag{17}$$

where

$$e^{N\Omega'_\mu}=\int\prod_j [P(\eta_j)d\eta_j]\exp\left(-\frac{1}{2}\sum_{i,j} G^0_{ij}\eta_i\eta_j\right), \tag{18}$$

$$\tau_j=\tau-\sum_i m_i G^0_{ij}, \tag{19}$$

$$P(\eta_j)=\frac{e^{\tau_j\eta_j}}{[e^{-\tau_j m_j}2\cosh(\tau_j)]}\{\delta(\eta_j-(1-m_j))$$
$$+\delta(\eta_j+(1+m_j))\}. \tag{20}$$

The m_j's are determined self-consistently from the condition

$$\langle\eta_j\rangle=\int d\eta_j P(\eta_j)\eta_j=0, \tag{21}$$

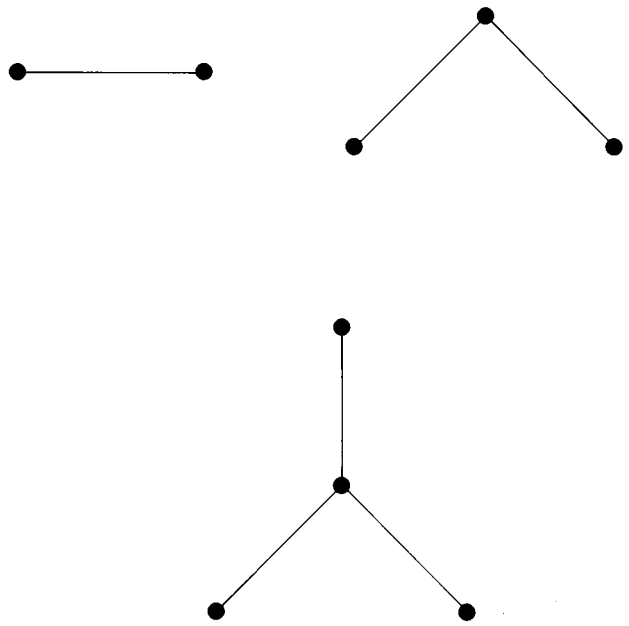

FIG. 2. Tree diagrams that vanish automatically when $b=0$.

which gives from Eq. (20)

$$m_j=\tanh(\tau_j). \tag{22}$$

The next task is the calculation of $N\Omega'_\mu$ in Eq. (18). We can use the cumulant expansion [7], to write

$$N\Omega'_\mu=\sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\left\langle\left[\sum_{(i,j)} G^0_{ij}\eta_i\eta_j\right]^k\right\rangle_c \tag{23}$$

where $\langle\ \rangle$ indicates an average over the probability distribution in Eq. (20).

$$\langle\ \rangle=\int\prod_j [P(\eta_j)d\eta_j](), \tag{24}$$

while $\langle\ \rangle_c$ means a cumulant average and $\Sigma_{(i,j)}$ a sum over bonds. The detailed analysis of the cumulant expansion in Eq. (23) is left for the Appendix, while we discuss here the main results.

$N\Omega'_\mu$ is given by the series of diagrams in Fig. 1, where a dot linking r legs at site j stands for a ''generalized cumulant'' average shown in Eq. (A7) for $r>1$, while a single line joining sites i and j stands for $(-1)G^0_{ij}$. Now G^0_{ii} is allowed from Eq. (15). The condition $\langle\eta_j\rangle=0$ insures that the tree diagrams in Fig. 2 with renormalized vertices automatically vanish.

Following standard methods in many body theory [8] we may write a renormalized expansion by defining a full propagator G_{ij} shown in Fig. 3(a) that satisfies Dyson's equation

$$G_{ij}=G^0_{ij}+\sum_{k,l} G^0_{ik}\Sigma_{kl}G_{lj} \tag{25}$$

where the self-energy Σ_{ij} is given by the diagrams that cannot be separated into two parts by cutting an internal line shown in Fig. 3(b).

Now $N\Omega'_\mu$ can be expressed as a function [8] of the matrix propagator $\underline{\underline{G}}$ with elements G_{ij}

$$N\Omega'_\mu = -\tfrac{1}{2}\mathrm{Tr}\,\log[\underline{\underline{G}}_0^{-1}\underline{\underline{G}}] + \frac{1}{2}\mathrm{Tr}[\underline{\underline{G}}\,\underline{\underline{G}}_0^{-1}] - \tfrac{1}{2}N\Phi(\{\underline{\underline{G}}\}) \tag{26}$$

and $N\Phi(\{\underline{\underline{G}}\})$ is the sum of skeleton diagrams shown in Fig. 4, where a double line stands for the full propagator G_{ij} in Eq. (25). The self-energy Σ_{ij} is given by

$$\Sigma_{ij} = \frac{\delta N\Phi(\{\underline{\underline{G}}\})}{\delta G_{ij}} \tag{27}$$

and $N\Omega'_\mu$ satisfies the stationarity condition

$$\frac{\delta N\Omega'_\mu}{\delta G_{ij}} = 0 \tag{28}$$

when G_{ij} and Σ_{ij} are given by Eqs. (25) and (27), respectively. These equations are formally exact. Meaningful approximations that respect the stationarity principle are obtained by approximating $N\Phi$ in Fig. 4. We discuss here the results in the single site or Brout approximation [9], when $N\Phi$ is approximated by the family of single site ''star'' diagrams shown in Fig. 5, which we call $N\Phi_{ss}$.

The detailed calculation in the Appendix gives from Eq. (A11)

$$N\Phi_{ss} = \sum_k \left\{ (1+m_k^2)G_{kk} + \log\left[\frac{\cosh(\tau_k - 2m_k G_{kk})}{\cosh(\tau_k)}\right]\right\}. \tag{29}$$

By introducing Eqs. (26) and (29) into Eq. (17) we obtain

$$N\Omega_\mu^{ss} = \sum_j \{-\nu_j + \tau m_j - \tfrac{1}{2}[(1+m_j)ln(1+m_j) + (1-m_j)ln(1-m_j) + (1+m_j^2)G_{jj} + \log\cosh(\tau_j - 2m_j G_{jj})$$

$$-\log\cosh(\tau_j)]\} - \tfrac{1}{2}\sum_{i,j} G^0_{ij} m_i m_j + \tfrac{1}{2}\mathrm{Tr}[\underline{\underline{G}}_0^{-1}\underline{\underline{G}}] - \tfrac{1}{2}\mathrm{Tr}\,\log[\underline{\underline{G}}_0^{-1}\underline{\underline{G}}] - N\,\log\cosh(\tau). \tag{30}$$

The stationarity condition in Eq. (27) gives $\Sigma_{ij} = \delta_{ij}\Sigma_j$, where

$$\Sigma_j = \frac{\delta N\Phi_{ss}}{\delta G_{jj}} = (1+m_j^2) - 2m_j \tanh[\tau_j - 2m_j G_{jj}]. \tag{31}$$

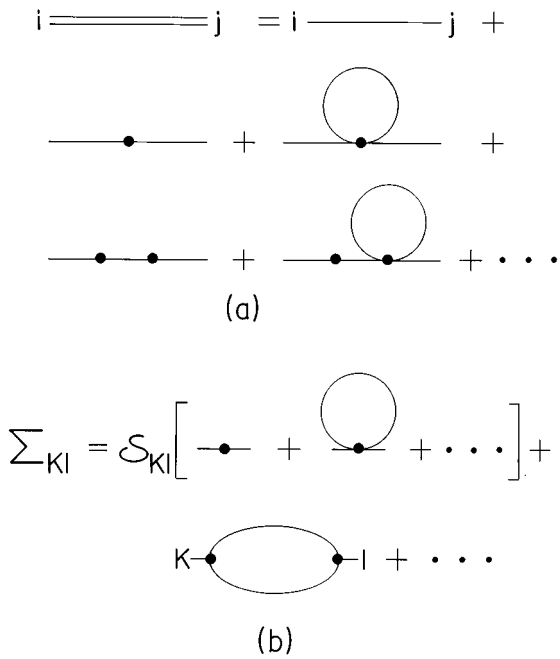

FIG. 3. (a) Diagrams for the full propagator G_{ij}. (b) Diagrams for the self-energy Σ_{kl}.

We also require $N\Omega_\mu(ss)$ to be stationary with respect to variations in m_j,

$$\frac{\delta N\Omega_\mu^{ss}}{\delta m_j} = (\nu_j - G_{jj})[m_j - \tanh(\tau_j - 2m_j G_{jj})] = 0 \tag{32}$$

and with respect to variations [6] in ν_j

$$\frac{\delta N\Omega_\mu^{ss}}{\delta \nu_j} = -1 + m_j^2 + \Sigma_{jj}\{[1 - \underline{\underline{G}}_0\underline{\underline{\Sigma}}]^{-1}\}_{jj} + m_j[\tanh(\tau_j) - \tanh(\tau_j - 2m_j G_{jj})] = 0. \tag{33}$$

From Eq. (32), m_j in Eq. (22) is a stationary point if

$$G_{jj} = \{[1 - \underline{\underline{G}}^0\underline{\underline{\Sigma}}]^{-1}\underline{\underline{G}}^0\}_{jj} = 0 \tag{34}$$

which gives from Eq. (31)

$$\Sigma_j = (1 - m_j^2), \tag{35}$$

and introducing Eq. (35) into Eq. (33) we obtain the equation for ν_j,

$$\{[1 - \underline{\underline{G}}^0\underline{\underline{\Sigma}}]^{-1}\}_{jj} = 1. \tag{36}$$

A more convenient form is derived from Eq. (34) with the help of Eqs. (15) and (36):

$$2\nu_j = \frac{i}{\sqrt{C}}\sum_k [1 - \underline{\underline{G}}^0\underline{\underline{\Sigma}}]^{-1}_{jk} t^\mu_{kj}. \tag{37}$$

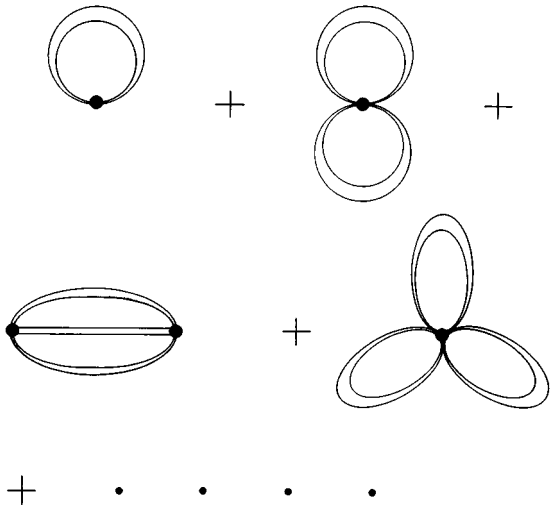

FIG. 4. Series of skeleton diagrams for $N\Phi$. Double lines indicate a full propagator G_{ij} as in Fig. 3(a).

If we expand Eq. (37) to second order in $(1/\sqrt{C})t_{ij}^{\mu}$ we obtain

$$2\nu_j \approx \sum_k \frac{i}{\sqrt{C}} t_{jk}^{\mu}(1-m_k^2)\frac{i}{\sqrt{C}} t_{kj}^{\mu} \tag{38}$$

which gives from Eq. (19) and Eq. (22);

$$m_j = \tanh\left[\tau - \frac{i}{\sqrt{C}}\sum_k m_k t_{kj}^{\mu} - m_j \sum_k \frac{i}{\sqrt{C}} t_{jk}^{\mu}(1-m_k^2)\frac{i}{\sqrt{C}} t_{kj}^{\mu}\right]. \tag{39}$$

What we obtain in Eq. (39) are TAP equations [5] for a spin glass with long range site dependent interactions $\beta\tilde{J}_{jk} = -(i/\sqrt{C})t_{jk}^{\mu}$.

By introducing Eqs. (34)–(36) into Eq. (30) and expanding $\log(\underline{\underline{G}}^{-1}\underline{\underline{G}}^{0})$ to second order in t_{kj}^{μ} as in Eq. (38) we obtain at the saddle point

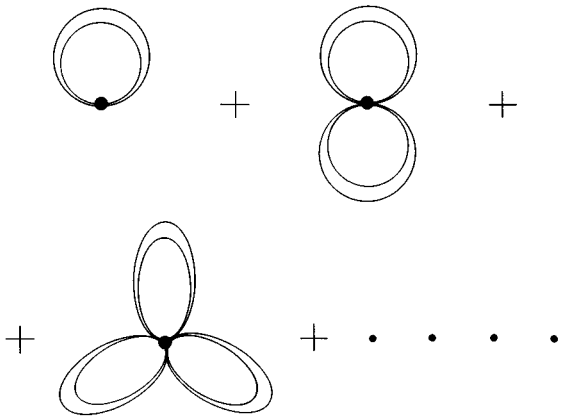

FIG. 5. Series of diagrams for $N\Phi_{ss}$ in the single site approximation.

$$N\Omega_{\mu}^{ss}]_{\mathrm{SP}} = \sum_j \left[\tau m_j - \tfrac{1}{2}(1+m_j)\ln(1+m_j) - \tfrac{1}{2}(1-m_j)\ln(1-m_j)\right] - \frac{i}{2\sqrt{C}}\sum_{k,j} t_{kj}^{\mu} m_k m_j - \frac{1}{4C}\sum_{k,j} t_{kj}^{\mu}(1-m_j^2)t_{jk}^{\mu}(1-m_k^2) - N\ \log\cosh(\tau), \tag{40}$$

which we recognize also as TAP free energy [5], which has an extremum when the m_j's are given by Eq. (39).

We went through a detailed and lengthy calculation of Ω_{μ} in order to exhibit the complexity and unsolubility of the problem. If we recall that from Eq. (13) the ''interactions'' t_{ij}^{μ} depend on the variables $x_{\mu i}^{\alpha}$ and J_{ij}^{α} that must be integrated in turn, we realize that it is hopeless to try to solve Eq. (39) exactly for nonvanishing values of the bias b.

A particularly simple result is obtained for *unbiased* patterns. By setting $b=\tau=0$ in Eq. (39) we see that it accepts the trivial, site independent solution $m_i=0$ which gives in Eq. (40)

$$N\Omega_{\mu}^{ss}(b=0)]_{\mathrm{SP}} = -\frac{1}{4C}\sum_{k,j} t_{kj}^{\mu} t_{jk}^{\mu}, \tag{41}$$

and introducing Eq. (41) into Eq. (11) we recover the result of Ref. [2]. This indicates that, within the range of validity of Eq. (39), the result in Ref. [2] is not restricted to extremely diluted systems. This conclusion deserves a more extended analysis of the contributing diagrams that we reserve for the last section.

The problem for biased patterns reaches another level of complexity. For $\tau \neq 0$, Eq. (39) does not accept a site independent solution; then the fundamental assumption of site independence made in Ref. [2] breaks down. Given the impossibility of obtaining a rigorous result, we present in the next section a simple approximation for $N\Omega_{\mu}^{ss}$ in Eq. (40), where we consider only tree diagrams with single unrenormalized bonds and sites as in Fig. 2 and the single unrenormalized bubble in Fig. 1(c). This approximation reproduces the results of Ref. [2] for $b=0, \eta \neq 0$ and those of Ref. [1] for $\eta = h = 0, b \ll 1$. The singular behavior of physical quantities when b approaches unity differs from the results in Ref. [1] because the problems are not exactly the same, even when $\eta = 0$.

IV. APPROXIMATION

In Eq. (39) we obtained a set of coupled nonlinear equations for the site dependent ''magnetizations'' or effective bias m_j. In order to decouple these equations we approximate the right-hand side of Eq. (39) by

$$m_j = \tanh\left[\tau - b\frac{i}{\sqrt{C}}\sum_k t_{kj}^{\mu}\right]. \tag{42}$$

In this way we break correlations by replacing the self-consistent field τ_j at every site by an effective field produced

by neurons with ''unperturbed'' bias b, hence we expect this to be a sensible approximation for small values of b.

Introducing Eq. (42) into Eq. (40) and approximating $m_k \approx b$ we obtain

$$N\Omega_\mu^{\rm eff} \approx \sum_j \ln\cosh\left[\tau - \frac{i}{\sqrt{C}} b \sum_k t_{kj}^\mu\right] - (1-b^2)^2 \frac{1}{4C}\sum_{k,j} (t_{kj}^\mu)^2 - N\ln\cosh(\tau). \quad (43)$$

As we discuss more extensively in the last section, this approximation amounts to sum tree diagrams with single unrenormalized bonds and vertices as those shown in Fig. 2, and to consider only the two sites unrenormalized bubble in Fig. 1(c).

The calculation of V_n in Eq. (11) follows according to standard procedures [1,2]. As in Ref. [1] we define the order parameter

$$M = \frac{1}{\sqrt{C}}\sum_k J_{ik}^\alpha \quad (44)$$

denominated ''ferromagnetic bias'' and as in Ref. [2] the order parameters for $\alpha \neq \beta$:

$$q = \frac{1}{C}\sum_k J_{ik}^\alpha J_{ik}^\beta, \qquad h = \frac{1}{C}\sum_k J_{ik}^\alpha J_{ki}^\beta. \quad (45)$$

The order parameters in Eqs. (44) and (45) are explicitly replica symmetric and site independent. We show in the Appendix the detailed calculations and mention here only the results. By introducing into Eq. (11) the order parameters in Eqs. (44) and (45) by means of a δ-function representation, together with Eq. (43), we obtain for $\log V$ in the limit $n=0$

$$\log V = \frac{NC}{2}\left\{\frac{1-\eta x}{(1-q)(1-x^2)} + \tfrac{1}{2}\ln[(1-q)^2(1-x^2)]\right\} + \alpha\left[-\frac{r^2}{(1-q)x} + \sum_{s=\pm 1}\frac{(1+sb)}{2}\int Dz\ \ln H(\chi_s)\right], \quad (46)$$

where

$$x = \frac{\eta - h}{1-q}, \quad (47)$$

$$\chi_s = \frac{1}{\sqrt{1-q}}\left\{\frac{1}{(1-b^2)}[\kappa - sMb(1+sb)] - r + z\sqrt{q}\right\}, \quad (48)$$

$$r = x[B_+ + B_-], \quad (49)$$

where we introduced the notation for $s = \pm 1$

$$B_s = \frac{1+sb}{2}\int Dz \frac{e^{-(1/2)\chi_s^2}}{\sqrt{2\pi}H(\chi_s)}\sqrt{1-q}, \quad (50)$$

$$Dz = \frac{dz}{\sqrt{2\pi}}e^{-(1/2)z^2}, \quad (51)$$

$$H(x) = \int_x^\infty Dz. \quad (52)$$

The expression for $\log V$ in Eq. (46) reproduces the result of Ref. [1] when $\eta = r = x = 0$, except for the expression for χ_s that coincides with Gardner's only for $b \ll 1$. It reproduces the result of Ref. [2] for arbitrary symmetry parameter η when $b=0$, hence we consider it to be a sensible mean field approximation for the volume of the space of interactions with arbitrary symmetry that have biased patterns as fixed points of the dynamic equations.

The saddle point equations for q, x, and M are obtained by extremizing $\log V$ in Eq. (46), and they are

$$\frac{q - 2hx + qx^2}{[1-x^2]^2} = \alpha\sum_{s=\pm 1}\frac{(1+sb)}{2}\int Dz\left[\frac{e^{-(1/2)\chi_s^2}}{\sqrt{2\pi}H(\chi_s)}\sqrt{1-q}\right]^2, \quad (53)$$

$$\frac{h - 2x + hx^2}{[1-x^2]^2} = \alpha\frac{r^2}{x^2}, \quad (54)$$

while we get for M

$$B_+ = B_-. \quad (55)$$

Due to the particular scaling in Eq. (44) for the ferromagnetic bias, we obtained also as in Ref. [1] that the volume depends on M only through χ_s in Eq. (48), and Eq. (55) is trivially satisfied when $b=0$. In this limit Eqs. (53) and (54) reduce to the equations of Gardner, Gutfreund, and Yekutieli in Ref. [2].

The critical storage capacity is obtained when $q=1$ and $h=\eta$, keeping x in Eq. (47) finite. We obtained from Eqs. (53) and (54) two coupled equations for α_c and x_c, where the value of M_c is obtained from Eq. (55). We show in Figs. 6 and 7 the results for α_c and M for $\kappa = 0$ as a function of the bias b for different values of the symmetry parameter η.

When b is close to unity the equations can be solved asymptotically and we obtain

$$M \approx \sqrt{-(1-b^2)^2\ln(1-b)}, \quad (56)$$

$$\alpha_c \approx \frac{[\sqrt{1+\eta} + \sqrt{1-\eta}]^2}{4M}. \quad (57)$$

V. CONCLUSIONS

The object of this paper is to analyze the volume of the space of interactions in a neural network satisfying a definite symmetry constraint as in Eq. (7) and accepting a configuration of *biased* patterns as fixed point of the dynamic equation.

The problem for asymmetric interactions $J_{ij} \neq J_{ji}$ was solved by Gardner [1], while the case of definite symmetry was studied by Gardner, Gutfreund, and Yekutieli only for *unbiased* patterns [2].

The main result of the present paper is that the bias in the patterns acts as an internal field and introduces correlations that map the problem into the spin-glass problem, thus preventing its solution. We show in Sec. III by means of rigorous diagrammatic expansions that to each site is associated an order parameter or ''effective bias'' m_j that satisfies TAP-like equations [5] for a spin glass. Just like in the spin-glass case the solution to these equations, if it could be found, is explicitly site dependent. However, for vanishing bias our equations accept the trivial solution $m_j = 0$ for all j that is explicitly site independent, and the results of our theory coincide with those of Ref. [2]. This seems highly surprising as the authors of this paper assert that their theory is only valid for extremely dilute networks with $C \ll \ln(N)$, while we have not made that assumption here. In order to understand this result we present a detailed analysis of the large N behavior of the unrenormalized diagrams that contribute to the TAP-like expression for $N\Omega_{\mu}^{ss}$ in Eq. (40), which should scale like N in the thermodynamic limit.

A. Tree diagrams

The first term in brackets in Eq. (40) sums the unrenormalized tree diagrams [4,7] shown in Fig. 2, where now each bare bond represents $it_{jk}^{\mu}/\sqrt{C}$.

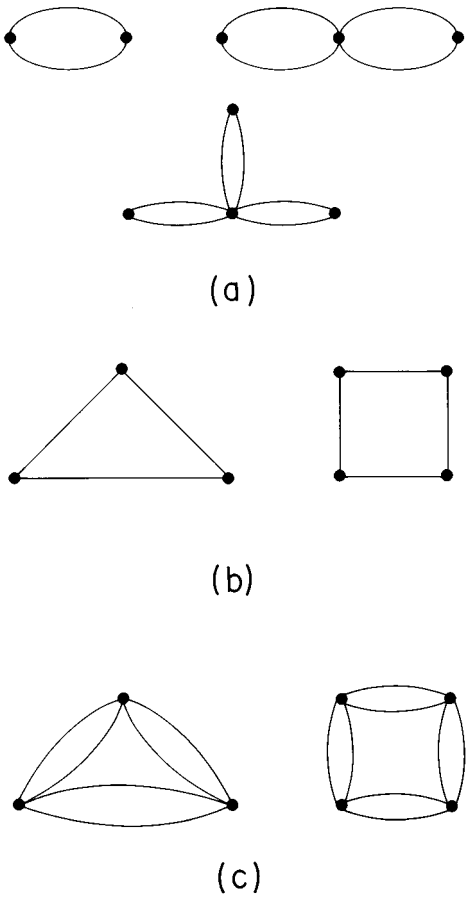

FIG. 6. (a) Tree of bubble diagrams. (b) Ring diagrams. (c) Double ring diagrams.

A tree diagram with s sites and $s-1$ bonds gives a contribution

$$T(s) \approx \left[\frac{C}{N}\right]^{s-1} \left[\frac{1}{\sqrt{C}}\right]^{s-1} \sum_{j} \left(\sum_{k} t_{kj}^{\mu} \right)^{s-1}, \tag{58}$$

where from Eq. (5) we introduced a factor C/N for each occupied bond. From Eq. (13) we obtain

$$\frac{1}{\sqrt{C}} \sum_{k} t_{kj}^{\mu} \approx \sum_{\alpha} x_{\mu j}^{\alpha} \frac{1}{\sqrt{C}} \sum_{k} J_{jk}^{\alpha} \approx M, \tag{59}$$

where M is the ferromagnetic bias in Eq. (44).

Introducing Eq. (59) into Eq. (58) gives $T(s) \approx (C/N)^{s-1} M^{s-1} N$, then the tree diagrams give the correct behavior in the thermodynamic limit $N \to \infty$ for the fully connected network with $C = N$.

B. Tree of bubbles

The single site diagrams in Fig. 5 together with the corresponding diagonal self-energy diagrams in Fig. 3 generate the *tree of bubbles* shown in Fig. 6(a). A typical diagram with s sites and $s-1$ doubly occupied bonds or *bubbles* gives a contribution

$$B(s) \approx \left(\frac{C}{N}\right)^{s-1} \sum_{j} \left[\frac{1}{C} \sum_{k} (t_{jk}^{\mu})^2 \right]^{s-1} \approx N, \tag{60}$$

where we used the scaling in Eq. (45) and they are well scaled for any $C \leqslant N$. All these diagrams except the single bubble with two sites vanish in the unbiased case, when $b = 0$. We conclude that within the single site approximation we are summing diagrams that are well scaled in the thermodynamic limit in the fully connected lattice.

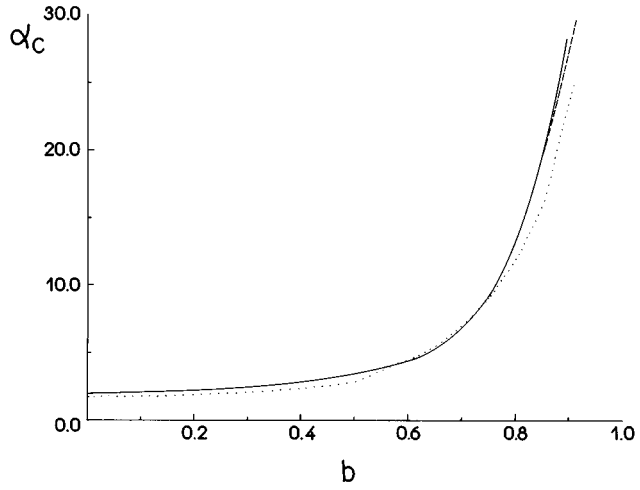

FIG. 7. Critical storing capacity α_c (dimensionless) as a function of the dimensionless bias b for different values of the asymmetry parameter η for the approximation in Eq. (42). Full line, broken line, and pointed line correspond to $\eta = 0.2$, $\eta = 0.4$, and $\eta = 0.6$, respectively.

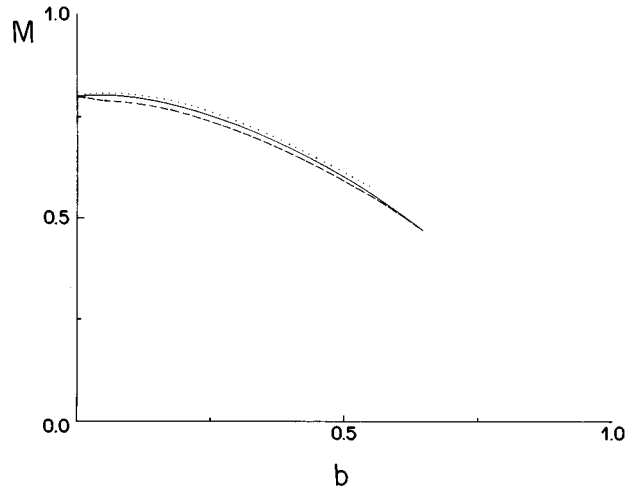

FIG. 8. Ferromagnetic bias M (dimensionless) as a function of the bias b for different values of η. Specifications are as in Fig. 7.

C. Ring diagrams

In order to derive Eq. (40) we expanded the saddle point equation for ν_j, Eq. (37), and $\mathrm{Tr}\log[\underline{\underline{G}}_0^{-1}\underline{\underline{G}}]$ in Eq. (30) to second order in t_{jk}^{μ}, which amounts to neglecting the ring diagrams with more than two sites that are shown in Fig. 6(b). What is the error that we are committing with this approximation? The estimate in Ref. [2] is that a ring diagram with s sites and s bonds is $O(C^{s/2})$, thus diverging in a connected lattice. It is claimed that the neglect of ring diagrams with $s \geqslant 3$ is only meaningful in the extremely dilute limit $C \ll \ln(N)$. The counting goes as follows: a ring diagram with s sites and s bonds contribute

$$R(s)=\left(\frac{C}{N}\right)^{s}\frac{1}{C^{s/2}}\sum_{i_1} t_{i_s i_1}^{\mu}\sum_{i_2} t_{i_1 i_2}^{\mu}\cdots\sum_{i_s} t_{i_{s-1} i_s}^{\mu}, \quad (61)$$

where the sum is restricted to $i_1 \neq i_2 \neq i_3 \ldots \neq i_s$. The pessimistic evaluation of Ref. [2] says that each summation gives a factor $O(N)$, then as a result $R(s)$ would be $O(C^{s/2})$, from where stems their assertion that the theory is only valid for extreme dilution. But if their counting were correct, the ferromagnetic bias in Eq. (44) would be $O(\sqrt{C})$, while our results in Fig. 8 give a finite value for M in agreement with the results in Ref. [1], thus showing the correct scaling.

A lower bound to $R(s)$ in Eq. (61) can be obtained by decoupling the sums and approximating

$$R(s)\approx\left(\frac{C}{N}\right)^{s}\left[\frac{1}{\sqrt{C}}\sum_{j} t_{jk}^{\mu}\right]^{s}\approx\left(\frac{C}{N}M\right)^{s}, \quad (62)$$

where we used Eq. (44). In this case we obtain that $R(s)\approx O(1)$ for $s \geqslant 3$ and it would be negligible, not divergent, in the thermodynamic limit. The ring diagram with $s=2$ scales differently and it gives the first contribution to the bubbles series. There are, however, two related problems where these diagrams can be calculated and give a meaningful answer for a fully connected lattice: the spin-glass problem discussed in Ref. [4] and the Hopfield model for a neural network with hierarchical clustering in Ref. [10], where we showed in the Appendix how the ring diagrams can be exactly summed. This happens because in all these well behaved cases the interactions J_{ij} take randomly positive and negative values, giving rise to cancellations, such as it happens in the present problem to have a finite value of M in Eq. (44).

Finally, we analyze an approximate solution where the effective bias m_j at each site is calculated by approximating $m_k \approx b, k \neq j$, in the local field at site j. This approximation amounts to sum only tree diagrams with single unrenormalized bonds and vertices and it interpolates between the results of Ref. [1] when $\eta=0, b\neq 0$ and those of Ref. [2] when $\eta\neq 0, b=0$. The results shown in Figs. 7 and 8 for the critical values of the storing capacity α_c and the ferromagnetic bias M exhibit small deviations with the values of the symmetry parameter η.

ACKNOWLEDGMENTS

Part of this work was performed during a visit to the Institut für Theoretische Physik, Universität Heidelberg. It is a pleasure to thank Professor H. Horner for the hospitality and for very fruitful interactions. Financial support from Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and Deutscher Akademischer Austauschdienst (DAAD) is kindly acknowledged.

APPENDIX

We present here a detailed diagrammatic analysis of $N\Omega'_{\mu}$ in Eq. (18). Following Horwitz and Callen [7] we can write

$$N\Omega'_{\mu}=\sum_{k=1}^{\infty}(-1)^{k}S_k[0], \quad (A1)$$

with

$$S_k[\gamma]=\sum_{(ij)}{}'\prod_{(ij)}\frac{1}{P_{ij}!}\left(G_{ij}^{0}\frac{\partial}{\partial\gamma_{ij}}\right)^{P_{ij}}\ln\left\langle\exp\left(\sum_{(ij)}\gamma_{ij}\eta_i\eta_j\right)\right\rangle, \quad (A2)$$

where $\Sigma_{(ij)}$ indicates a sum over bonds and P_{ij} are all positive integers subject to the restriction $\Sigma_{(ij)}P_{ij}=k$. The average over the η_j's is to be taken with the probability distribution in Eq. (20), that gives for the first three moments

$$\langle\eta_j\rangle=0,$$
$$\langle\eta_j^2\rangle=1-m_j^2, \quad (A3)$$
$$\langle\eta_j^3\rangle=-2m_j(1-m_j^2),$$

etc., and we notice that $\langle\eta_j^r\rangle\neq 0$ for $r>1$. We obtain for the first terms in the series in Eq. (A1) recalling that $G_{ii}^{0}\neq 0$

$$S_1(0)=\frac{1}{2}\sum_{i} G_{ii}^{0}\langle\eta_i^2\rangle, \quad (A4)$$

$$S_2(0)=\frac{1}{4}\sum_i (G^0_{ii})^2\langle[\eta_i^2]^2\rangle_{\rm gc}+\frac{1}{4}\sum_{i\neq j}(G^0_{ij})^2\langle\eta_i^2\rangle\langle\eta_j^2\rangle, \tag{A5}$$

$$\begin{aligned}S_3(0)=&\frac{1}{12}\sum_i (G^0_{ii})^3\langle[\eta_i^2]^3\rangle_{\rm gc}+\frac{1}{12}\sum_{i\neq j}(G^0_{ij})^3\langle\eta_i^3\rangle\langle\eta_j^3\rangle\\&+3\sum_{j,k} G^0_{jj}(G^0_{jk})^2\langle[\eta_j^2]^2\rangle_{\rm gc}\langle\eta_k^2\rangle\\&+\frac{1}{6}\sum_{i\neq j=k} G^0_{ij}G^0_{jk}G^0_{ki}\langle\eta_i^2\rangle\langle\eta_j^2\rangle\langle\eta_k^2\rangle,\end{aligned} \tag{A6}$$

where now the sums are over sites.

Each of the terms in Eq. (A4) to Eq. (A6) may be represented by a diagram as shown in Fig. 1, where a single line between sites i and j represents G^0_{ij}, a dot at site j where m ''ears'' enter (or closed lines) and n lines stands for $\langle[\eta_j^2]^m[\eta_j^n]\rangle_{\rm gc}$, and the first ''generalized cumulants'' are given by

$$\begin{aligned}\langle[\eta_j^2]\rangle_{\rm gc}&=\langle\eta_j^2\rangle,\\ \langle[\eta_j^2]^2\rangle_{\rm gc}&=\langle\eta_j^4\rangle-\langle\eta_j^2\rangle^2,\\ \langle[\eta_j^2]^3\rangle_{\rm gc}&=\langle\eta_j^6\rangle-3\langle\eta_j^4\rangle\langle\eta_j^2\rangle+2\langle\eta_j^2\rangle^3,\\ \langle[\eta_j^2][\eta_j^3]\rangle_{\rm gc}&=\langle\eta_j^5\rangle-\langle\eta_j^2\rangle\langle\eta_j^3\rangle,\end{aligned} \tag{A7}$$

etc. Then $S_1(0)$ is represented in Fig. 1(a), the two terms for $S_2(0)$ in Figs. 1(b) and 1(c), to $S_3(0)$ corresponds to Figs. 1(d)–1(g), etc.

The diagrammatic series in Fig. 1 may be renormalized by introducing the ''propagator''

$$G_{ij}=G^0_{ij}+\sum_{k,l} G^0_{ik}\Sigma_{kl}G_{lj} \tag{A8}$$

represented in Fig. 3(a), where the ''self-energy'' Σ_{kl} is shown in Fig. 3(b). Following standard methods in many body theory [8], the quantity $N\Omega'_\mu$ can be written as the stationary functional of G_{ij} shown in Eq. (26).

We now calculate $N\Phi_{ss}$, the sum of single site ''star'' skeleton diagrams shown in Fig. 5:

$$N\Phi_{ss}=\sum_j \Phi_j, \tag{A9}$$

where, from Eq. (A4) to Eq. (A7),

$$\Phi_j=\sum_{n=1}^{\infty}\frac{1}{n!}\langle[\eta^2]^n\rangle_{\rm gc}(G_{jj})^n=\log\langle e^{\eta_j^2 G_{jj}}\rangle \tag{A10}$$

and we obtain from Eq. (20)

$$\langle e^{\eta_j^2 G_{jj}}\rangle=e^{(1+m_j^2)G_{jj}}\frac{\cosh(\tau_j-2m_jG_{jj})}{\cosh(\tau_j)}. \tag{A11}$$

Introducing Eq. (A11) into Eq. (A10) we obtain Eq. (29).

Calculation of the volume in the effective field approximation

In order to calculate the volume in phase space within the effective field approximation in Sec. IV we introduce integral representations for the constraints δ functions and for the relations defining the order parameters in Eqs. (44) and (45). We assume here replica symmetry and site independence [2], but it is understood that the validity of this assumption is limited to the effective field approximation in Sec. IV and it does not hold for the general results in Sec. III when $b\neq 0$. By introducing Eq. (43) into Eq. (11) we obtain

$$\begin{aligned}\log V=-NC\Bigg\{&\frac{1}{2}\left[E+\eta G-Fq-Hh+\frac{MP}{\sqrt{C}}\right]\\&+\frac{1}{4}\ln[(G-H)^2-(E-F)^2]+\frac{i}{4}\frac{P^2}{E-F+(G-H)}\\&+\frac{1}{4}\frac{F+H}{E-F+(G-H)}+\frac{1}{4}\frac{F-H}{E-F-(G-H)}-\alpha Q\Bigg\}_{\rm SP},\end{aligned} \tag{A12}$$

where the variables should be taken at their saddle point values. In Eq. (A12) Q is given by

$$\begin{aligned}Q=\Bigg\{&-i[\varphi\Phi+\psi\Psi+wW]+\sum_{s=\pm 1}\tfrac{1}{2}(1+sb)\int Dz\ \ln H(\tilde{\chi}_s)\\&-\tfrac{1}{4}(1-b^2)^2[(1-q)\psi+qw]-iMb^2\varphi\Bigg\}_{\rm SP},\end{aligned} \tag{A13}$$

where

$$\begin{aligned}\tilde{\chi}_s=\Bigg(&\kappa+\Phi-Mbs+i\frac{\varphi}{2}(\eta-h)(1-b^2)^2\\&+z\left[-2iW+\frac{1}{2}q(1-b^2)^2\right]^{1/2}\Bigg)\\&\times\frac{1}{[-2i\Psi+\frac{1}{2}(1-q)(1-b^2)^2]^{1/2}}\end{aligned} \tag{A14}$$

and the variables $\varphi,\Phi,\psi,\Psi,w,V$ should be taken at their saddle point value that extremizes Q. We obtain at the saddle point

$$\begin{aligned}\Phi&=\frac{i}{2}\varphi(\eta-h)(1-b^2)^2-iMb^2,\\ \Psi&=\frac{i}{4}(1-q)(1-b^2)^2,\\ W&=\frac{i}{4}q(1-b^2)^2,\\ -i\varphi&=\left[\sum_{s=\pm 1}\frac{(1+sb)}{2}\int Dz\frac{e^{-(1/2)\chi_s^2}}{\sqrt{2\pi}H(\chi_s)}\right]\\&\quad\times\frac{1}{[-2i\Psi+\frac{1}{2}(1-q)(1-b^2)^2]^{1/2}},\end{aligned} \tag{A15, A16}$$

while the expressions for w and ψ are immaterial because they drop from Q at the saddle point. We obtain from Eqs. (A13) and (A14)

$$Q=-\frac{1}{2}\left\{\frac{r^2}{\eta-h}-\sum_{s=\pm 1}(1+sb)\int Dz\ \ln(\chi_s)\right\}, \tag{A17}$$

$$\chi_s=\frac{1}{\sqrt{1-q}}\left\{\frac{\kappa-Mbs(1+bs)}{(1-b^2)}-r+z\sqrt{q}\right\}, \tag{A18}$$

where we called $r=-i\varphi(\eta-h)(1-b^2)$ and from Eq. (A16)

$$r=\frac{(\eta-h)}{\sqrt{1-q}}\sum_s\frac{(1+sb)}{2}\int Dz\frac{e^{-(1/2)\chi_s^2}}{\sqrt{2\pi}H(\chi_s)}. \tag{A19}$$

We observe that, just like in Gardner's work [1], the auxiliary field P associated with the ferromagnetic bias M appears multiplied by $C^{-1/2}$ and drops out in the thermodynamic limit. By solving the saddle point equations for E,F,G,H, we can recast $\log V$ in Eq. (A12) into Eq. (46) in the main text.

[1] E. Gardner, J. Phys. A **21**, 257 (1988).
[2] E. Gardner, H. Gutfreund, and I. Yekutieli, J. Phys. A **22**, 1995 (1989).
[3] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975); K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).
[4] H. J. Sommers, Z. Phys. B **31**, 301 (1978).
[5] D. J. Thouless, P. W. Anderson, and R. G. Palmer, Philos. Mag. **35**, 593 (1977).
[6] B. Southern and A. P. Young, J. Phys. C **10**, L79 (1977).
[7] G. Horwitz and H. B. Callen, Phys. Rev. **124**, 1757 (1961).
[8] See, for instance, A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinsky, *Methods of Quantum Field Theory in Statistical Physics* (Prentice Hall, Englewood Cliffs, NJ, 1963).
[9] R. Brout, Phys. Rev. **122**, 469 (1961).
[10] M. A. P. Idiart and Alba Theumann, J. Phys. A **25**, 779 (1992).