Order-parameter flow in symmetric and nonsymmetric fully connected attractor neural networks near saturation

S. N. Laughton and A. C. C. Coolen
Department of Physics—Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, United Kingdom
(Received 28 September 1994)

We apply a recent theory by Coolen and Sherrington [Phys. Rev. E. **49**, 1921 (1994)] which describes the dynamics of the Hopfield model near saturation in terms of deterministic flow equations for order parameters to the more general and technically more complicated case of neural networks with (i) arbitrary separable interactions, which (ii) need not be symmetric, and with (iii) more than one condensed pattern. Following the key assumptions of the previous theory, the distribution of intrinsic noise components of the alignment fields is calculated with the replica method. In the region where replica symmetry is stable, numerical simulations show that our equations capture the essential features of the flow, even for nonsymmetric systems (i.e., without detailed balance). For symmetric systems, the fixed points of the flow are shown to reproduce the thermodynamic equilibrium equations recently obtained by Cugliandolo and Tsodyks [J. Phys. A **27**, 741 (1994)].

PACS number(s): 05.50.+q, 05.70.−a, 87.10.+e

I. INTRODUCTION

Recently a method has been proposed for deriving a closed set of equations governing the evolution of macroscopic order parameters in the Hopfield [1] neural network model near saturation [2,3]. This method, based on the systematic removal of microscopic memory effects, has subsequently been applied to other disordered spins systems [4,5], and is understood to be exact at least (i) for short times (upon appropriate choice of initial conditions), (ii) in equilibrium, and (iii) in the limit where the disorder is removed (i.e., for attractor neural networks away from saturation). For an overview of the method and its present applications we refer to [6]. Although for intermediate time scales the procedure is not exact (manifested in an overall slowing down of the dynamics, which the theory does not account for), it does capture the essential characteristics of the flows and recovers the well known equilibrium properties of the archetypal disordered spin systems [7,8] as stable fixed points of the dynamic equations, including the full replica formalism. Furthermore, in contrast to phenomenological strategies for deriving flow equations, based on or inspired by time-dependent Landau-Ginzburg-type equations as in [9,10] or based on making an (incorrect) Gaussian ansatz for the local field distribution as in [11] (with various degrees of success in explaining dynamical phenomena), the present theory is derived from microscopic principles and generates explicitly the non-Gaussian shape of the local field distribution, which is ultimately responsible for the spin-glass-type features of the dynamics.

Analyzing the Hopfield model near saturation in its easiest form, within the so-called condensed ansatz and upon assuming only one condensed pattern, is already technically more involved than analyzing, for instance, the Sherrington-Kirkpatrick spin-glass model [5]. This is true within equilibrium statistical mechanics as well as from a dynamical point of view. Yet the Hopfield model is in fact the simplest fully connected attractor neural network with regard to the structure of the interactions between the neurons. It is the aim of the present paper to apply the theory in [2,3,6] to the more general and technically more complicated case of fully connected neural networks near saturation with (i) arbitrary separable interactions, which (ii) need not be symmetric, and with (iii) more than one condensed pattern. Since absence of symmetry of the neural interactions implies absence of detailed balance, our analysis includes systems for which equilibrium statistical mechanics does not apply, so that hitherto there has been no theory available with which they could be studied. This latter aspect was in fact one of the main motivations behind the development of the theory in [2,3].

Following the key assumptions of the theory in [2,3,6], self-averaging of the macroscopic flow with respect to the disorder and equipartitioning of probability within the macroscopic subshells of the ensemble, the distribution of intrinsic noise components of the alignment fields is calculated with the replica method. In our analysis we make the replica-symmetric ansatz. In the region where replica symmetry is stable, numerical simulations on large systems (40 000 spins), using the method proposed by Kohring [12], show that for the models considered our equations capture the essential features of the flow, even for nonsymmetric choices for the neural interactions (i.e., without detailed balance). One specific implication of this result is that we now have a general theory with which to calculate, for instance, the storage capacity of a class of attractor neural networks in which the stored attractors are limit cycles as opposed to fixed points. For symmetric systems, the fixed points of our flow equations are shown to reproduce the thermodynamic equilibrium equations recently obtained for the present class of networks with separable interactions by Cugliandolo and Tsodyks [13].

II. DYNAMICS OF SEPARABLE ATTRACTOR NEURAL NETWORKS NEAR SATURATION

We consider a system of N Ising spin neurons $s_i \in \{-1,1\}$, evolving in time according to a stochastic alignment to local fields h_i. This process is described by a master equation for the microscopic probability distribution $p_t(\mathbf{s})$:

$$\frac{dp_t(\mathbf{s})}{dt} = \sum_i \{w_i(F_i\mathbf{s})p_t(F_i\mathbf{s}) - w_i(\mathbf{s})p_t(\mathbf{s})\} , \tag{1}$$

where F_i is the spin flip operator

$$F_i\Phi(s_1,\ldots,s_i,\ldots,s_N) = \Phi(s_1,\ldots,-s_i,\ldots,s_N) ,$$

and where the transition rates $w_i(\mathbf{s})$ given by

$$w_i(\mathbf{s}) = \tfrac{1}{2}\{1 - s_i \tanh[\beta h_i(\mathbf{s})]\} ,$$

$$h_i = \sum_{j\neq i} J_{ij}s_j , \tag{2}$$

$$J_{ij} = \frac{1}{N}\sum_{\mu,\nu=1}^{p} \xi_i^\mu A_{\mu\nu}\xi_j^\nu ,$$

lead to the familiar Boltzmann form $p_\infty(\mathbf{s}) \propto e^{-\beta H(\mathbf{s})}$ in equilibrium for the special case where the interaction matrix $\{J_{ij}\}$ is symmetric. The "inverse temperature" $\beta = T^{-1}$ measures the stochasticity in the process. The interactions $\{J_{ij}\}$ are chosen to be of a general separable form, where the randomly drawn variables $\xi_i^\mu \in \{-1,+1\}$ represent bits of patterns embedded in the system by virtue of (2). For $A_{\mu\nu} = \delta_{\mu\nu}$ we recover the Hopfield model, the equilibrium properties of which were studied in the seminal papers [7,14] (including the model near saturation, i.e., $p = \alpha N$ with $\alpha > 0$). The dynamics, for arbitrary $\{A_{\mu\nu}\}$ but finite p, was analyzed in [15]. The equilibrium statistical mechanical analysis near saturation for symmetric $\{A_{\mu\nu}\}$ (detailed balance) has only recently been performed in [13]. Here we concentrate on the dynamics near saturation, i.e., $p = \alpha N$, or arbitrary $\{A_{\mu\nu}\}$ (not necessarily symmetric).

We make the so-called condensed ansatz: only a finite number c of patterns $\xi^\mu = (\xi_1^\mu,\ldots,\xi_N^\mu)$ are assumed to have a finite overlap with the system state $\mathbf{s}$; the remaining $p-c$ overlaps are assumed to be $O(1/\sqrt{N})$. The pattern components ξ_i^μ for $\mu > c$ are regarded as (quenched) disorder. We now choose as our dynamic order parameters the condensed overlaps $\mathbf{m}(\mathbf{s}) = (m^1(\mathbf{s}),\ldots,m^c(\mathbf{s}))$, and a state variable $r(\mathbf{s})$ to represent the uncondensed overlaps:

$$m^\mu(\mathbf{s}) = \frac{1}{N}\sum_i \xi_i^\mu s_i , \quad \mu = 1,\ldots,c , \tag{3}$$

$$\alpha r(\mathbf{s}) = \sum_{[\mu,\nu>c|\mu>c,\nu\le c|\mu\le c,\nu>c]} \left[\frac{1}{N}\sum_j \xi_j^\mu s_j\right] A_{\mu\nu} \left[\frac{1}{N}\sum_j \xi_j^\nu s_j\right] . \tag{4}$$

For symmetric systems $r(\mathbf{s})$ is proportional to the disorder-dependent contribution to the Hamiltonian (the disorder-independent contribution being a function of $\mathbf{m}$). As in [2,3,6] we hereby build in the correct equilibrium behavior. For nonsymmetric systems, however, no such guide for choosing r is available and the only motivations are analogy with the symmetric case and *(a posteriori)* success of the resulting theory. Since $r(\mathbf{s})$ contains a symmetric sum, it only depends on the symmetric part $\mathbf{A}^s$ of the matrix $\mathbf{A}$.

The corresponding macroscopic probability distribution $\mathcal{P}_t(\mathbf{m},r)$ is

$$\mathcal{P}_t(\mathbf{m},r) = \sum_{\mathbf{s}} p_t(\mathbf{s})\delta(\mathbf{m}-\mathbf{m}(s))\delta(r - r(\mathbf{s})) . \tag{5}$$

Insertion of the microscopic laws (1) leads to

$$\begin{aligned}\frac{d}{dt}\mathcal{P}_t(\mathbf{m},r) = {}& \frac{\partial}{\partial \mathbf{m}}\sum_{\mathbf{s}} p_t(\mathbf{s})\frac{1}{N}\sum_i \{1 - s_i\tanh[\beta h_i(\mathbf{s})]\}s_i\xi_i\delta(\mathbf{m}-\mathbf{m}(s))\delta(r-r(\mathbf{s})) \\ & + 2\frac{\partial}{\partial r}\sum_{\mathbf{s}} p_t(\mathbf{s})\frac{1}{p}\sum_i \{1 - s_i\tanh[\beta h_i(\mathbf{s})]\}s_i z_i^s(\mathbf{s})\delta(\mathbf{m}-\mathbf{m}(\mathbf{s}))\delta(r-r(\mathbf{s})) + O\left[\frac{c^2}{N}\right] ,\end{aligned} \tag{6}$$

where $\xi_i = (\xi_i^1,\ldots,\xi_i^c)$ and

$$h_i(\mathbf{s}) = \sum_{\mu,\nu\le c} \xi_i^\mu A_{\mu\nu} m^\nu(\mathbf{s}) + z_i^s(\mathbf{s}) + z_i^a(\mathbf{s}) ,$$

$$z_i^{s,a} = \sum_{[\mu,\nu>c|\mu>c,\nu\le c|\mu\le c,\nu>c]} \xi_i^\mu A_{\mu\nu}^{s,a}\left[\frac{1}{N}\sum_{j\neq i}\xi_j^\nu s_j\right] . \tag{7}$$

Using $(1/p)\sum_i s_i z_i^s = r(\mathbf{s}) - (1/p)\mathrm{Tr}\,\mathbf{A} + O(1/\sqrt{N})$, we can write (6) as

$$\frac{d}{dt}\mathcal{P}_t(\mathbf{m},r)=\nabla_m\cdot\left\{\mathcal{P}_t(\mathbf{m},r)\left[\mathbf{m}-\left\langle\frac{1}{N}\sum_i \xi_i \tanh\beta\left[\sum_{\mu,\nu\le c}\xi_i^\mu A_{\mu\nu}m^\nu+z_i^s+z_i^a\right]\right\rangle_{\mathbf{m},r;t}\right]\right\}$$
$$+2\frac{\partial}{\partial r}\left\{\mathcal{P}_t(\mathbf{m},r)\left[r-\frac{1}{p}\mathrm{Tr}\,\mathbf{A}-\frac{1}{\alpha}\left\langle\frac{1}{N}\sum_i z_i^s \tanh\beta\left[\sum_{\mu,\nu\le c}\xi_i^\mu A_{\mu\nu}m^\nu+z_i^s+z_i^a\right]\right\rangle_{\mathbf{m},r;t}\right]\right\}+O\left[\frac{c^2}{N}\right] \tag{8}$$

where we have introduced the subshell average

$$\langle\Phi\rangle_{\mathbf{m},r;t}=\frac{\sum_{\mathbf{s}}p_t(\mathbf{s})\delta(\mathbf{m}-\mathbf{m}(\mathbf{s}))\delta(r-r(\mathbf{s}))\Phi(\mathbf{s})}{\sum_{\mathbf{s}}p_t(\mathbf{s})\delta(\mathbf{m}-\mathbf{m}(\mathbf{s}))\delta(r-r(\mathbf{s}))} . \tag{9}$$

In the limit $N\to\infty$ Eq. (8) acquires the Liouville form, the solutions of which describe deterministic evolution, $\mathcal{P}_t(\mathbf{m},r)=\delta(\mathbf{m}-\mathbf{m}^*(t))\delta(r-r^*(t))$, where $(m^*(t),r^*(t))$ obeys the flow equations

$$\frac{d}{dt}\mathbf{m}=\int dz^s dz^a\langle\mathcal{D}_\xi[z^s,z^a]\xi\tanh\beta(\xi\cdot\mathbf{A}_{cc}\mathbf{m}+z^s+z^a)\rangle_\xi-\mathbf{m} , \tag{10}$$

$$\frac{1}{2}\frac{d}{dt}r=\frac{1}{\alpha}\int dz^s dz^a\langle\mathcal{D}_\xi[z^s,z^a]z^s\tanh\beta(\xi\cdot\mathbf{A}_{cc}\mathbf{m}+z^s+z^a)\rangle_\xi-r+\frac{1}{p}\mathrm{Tr}\,\mathbf{A} , \tag{11}$$

with $\langle f(\xi)\rangle_\xi=2^{-c}\sum_{\xi\in\{-1,1\}^c}f(\xi)$, and with the distributions $\mathcal{D}_\xi[z^s,z^a]$ of the intrinsic noise contributions to the alignment fields within sublattices,

$$\mathcal{D}_\xi[z^s,z^a]=2^c\left\langle\frac{1}{N}\sum_i\delta(z^s-z_i^s(\mathbf{s}))\delta(z^a-z_i^a(\mathbf{s}))\delta_{\xi\xi_i}\right\rangle_{\mathbf{m},r;t} , \quad \int dz^s dz^a\langle\mathcal{D}_\xi[z^s,z^a]\rangle_\xi=1 . \tag{12}$$

So far the theory is exact for $N\to\infty$, within the condensed ansatz. However, the noise distributions depend on t through the microscopic probability distribution $p_t(\mathbf{s})$, requiring us to solve the master equation (1), which is exactly what we want to avoid. We now close the macroscopic equations [(10) and (11)], following [2,3,6], by assuming the following:

(i) The flow equations [(10) and (11)], and hence the noise distributions, are self-averaging with respect to the disorder, i.e., the variables ξ_i^μ, allowing us to average over them.

(ii) We can assume equipartitioning of probability within the $(\mathbf{m},r)$ subshells of the ensemble as far as the calculation of the $\mathcal{D}_\xi[z^s,z^a]$ is concerned.

N.B. we do not assume that the microscopic probability distribution $p_t(\mathbf{s})$ obeys such equipartitioning, rather that probability fluctuations within the subshells are negligible once the average over patterns has been carried out. These two assumptions, the correctness of which can be verified only by comparison of the predictions of the resulting theory with numerical simulations, close the flow equations (11) and reduce our problem to that of calculating

$$\mathcal{D}_\xi[z^s,z^a]=\left\langle\frac{\sum_{\mathbf{s}}\delta(\mathbf{m}-\mathbf{m}(\mathbf{s}))\delta(r-r(\mathbf{s}))\delta(z^s-z_i^s(\mathbf{s}))\delta(z^a-z_i^a(\mathbf{s}))}{\sum_s\delta(\mathbf{m}-\mathbf{m}(\mathbf{s}))\delta(r-r(\mathbf{s}))}\right\rangle_\eta \quad \text{for } \xi_i=\xi . \tag{13}$$

Here we have redefined the pattern components to be averaged over to differentiate them from the condensed patterns: $\eta_j^\mu=\xi_j^\mu$ for $\mu>c$.

III. INTRINSIC NOISE DISTRIBUTIONS

A. Replica approach

We calculate the noise distributions using the replica identity

$$\frac{\langle\Phi(\mathbf{s})W(\mathbf{s})\rangle_s}{\langle W(\mathbf{s})\rangle_s}\equiv\lim_{n\to 0}\left\langle\Phi(\mathbf{s}^1)\prod_{\alpha=1}^n W(\mathbf{s}^\alpha)\right\rangle_{\{\mathbf{s}^\alpha\}} \tag{14}$$

Upon insertion of integral representations for the various δ distributions we can write the intrinsic noise distributions as

$$\mathcal{D}_{\zeta}[z^s,z^a]=\lim_{n\to 0}\left[\frac{N}{2\pi}\right]^{2n}\int dx\,dy\,e^{ixz^s+iyz^a}\int\left[\prod_{\alpha} d\mathbf{M}^{\alpha}dR^{\alpha}\right]e^{iN\sum_{\alpha}(\mathbf{M}^{\alpha}\cdot\mathbf{m}+R^{\alpha}r)}$$
$$\times\left\langle\left\langle \exp[-ixz^s(\mathbf{s}^1)-iyz^a(\mathbf{s}^1)-iN\sum_{\alpha}(\mathbf{M}^{\alpha}\cdot\mathbf{m}(\mathbf{s}^{\alpha})+R^{\alpha}r(\mathbf{s}^{\alpha}))]\right\rangle_{\eta'}\right\rangle_{\{\mathbf{s}^{\alpha}\}} \tag{15}$$

with the (Greek) replica indices $\alpha=1,\dots,n$. We now define $(p-c)n$ quantities $z^{\mu}_{\alpha}=(1/\sqrt{N})\sum_{k\neq i}\eta^{\mu}_k s^{\alpha}_k\sim O(1)$, $(\mu>c)$, so that the averages in the above expression becomes

$$\left\langle\left\langle \exp\left[-i\left\{\sum_{\mu>c}\eta^{\mu}_i\left[\sum_{\nu>c}(xA^s_{\mu\nu}+yA^a_{\mu\nu})\frac{z^{\nu}_1}{\sqrt{N}}+\sum_{\nu\le c}(xA^s_{\mu\nu}+yA^a_{\mu\nu})m^{\nu}\right]+\sum_{\mu\le c,\nu>c}\zeta^{\mu}(xA^s_{\mu\nu}+yA^a_{\mu\nu})\frac{z^{\nu}_1}{\sqrt{N}}\right\}\right]\right.\right.$$
$$\times\exp\left[-iN\sum_{\alpha}\left\{\mathbf{M}^{\alpha}\cdot\mathbf{m}(\mathbf{s}^{\alpha})+\frac{R^{\alpha}}{\alpha N}\left[\sum_{\mu,\nu>c}\left[z^{\mu}_{\alpha}+\frac{\eta^{\mu}_i s^{\alpha}_i}{\sqrt{N}}\right]A^s_{\mu\nu}\left[z^{\nu}_{\alpha}+\frac{\eta^{\nu}_i s^{\alpha}_i}{\sqrt{N}}\right]\right.\right.\right.$$
$$\left.\left.\left.\left.\left.+2\sqrt{N}\sum_{\mu>c,\nu\le c}\left[z^{\mu}_{\alpha}+\frac{\eta^{\mu}_i s^{\alpha}_i}{\sqrt{N}}\right]A^s_{\mu\nu}m^{\nu}\right]\right\}\right]\right\rangle_{\eta'}\right\rangle_{\{s^{\alpha}\}} \tag{16}$$

Assuming that $\sum_{\nu>c}\eta^{\mu}_i A^s_{\mu\nu}\eta^{\nu}_i\sim O(1)$, so that this term can be neglected, we can average over $\{\eta^{\mu}_i\}$ and $\{z^{\mu}_i\}$ upon introduction of the density of states

$$\mathcal{D}(\mathbf{z},\mathbf{s})=\left\langle\delta\left[\mathbf{z}-\frac{1}{\sqrt{N}}\sum_{k\neq i}\boldsymbol{\eta}_k\otimes\mathbf{s}_k\right]\right\rangle_{\eta}=\frac{e^{-(1/2)\mathbf{z}\cdot(\mathbf{q}^{-1}\otimes\mathbb{1})\mathbf{z}}}{\sqrt{(2\pi)^{n(p-c)}\det(\mathbf{q}\otimes\mathbb{1})}}\quad(\text{for }N\to\infty)\ , \tag{17}$$

where $q_{\alpha\beta}(\mathbf{s})=(1/N)\sum_{k\neq i}s^{\alpha}_k s^{\beta}_k$, and $\mathbb{1}_{\mu\nu}=\delta_{\mu\nu}$. Our average (16) therefore becomes

$$\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{s})\exp\left[-\frac{i}{\alpha}\sum_{\alpha}R^{\alpha}\mathbf{z}_{\alpha}\cdot\mathbf{A}^s_{uu}\mathbf{z}_{\alpha}\right]\exp\left[-i\sum_{\alpha}\sum_{\mu\le c,\nu>c}\left[\zeta^{\mu}(xA^s_{\mu\nu}+yA^a_{\mu\nu})\frac{\delta_{\alpha 1}}{\sqrt{N}}+\frac{2}{\alpha}\sqrt{N}R^{\alpha}m^{\mu}A^s_{\mu\nu}\right]z^{\nu}_{\alpha}\right]$$
$$\times\prod_{\mu>c}\cos\left[\left[\sum_{\nu>c}(xA^s_{\mu\nu}+yA^a_{\mu\nu})\frac{z^{\nu}_1}{\sqrt{N}}+\sum_{\nu\le c}(xA^s_{\mu\nu}+yA^a_{\mu\nu})m^{\nu}\right]+\frac{2}{\alpha}\sum_{\alpha}R^{\alpha}\left[\sum_{\nu>c}A^s_{\mu\nu}\frac{s^{\alpha}_i z^{\nu}_{\alpha}}{\sqrt{N}}+\sum_{\nu\le c}A^s_{\mu\nu}s^{\alpha}_i m^{\nu}\right]\right]\ . \tag{18}$$

In order to evaluate the integral we expand $\cos x=1-x^2/2+O(x^4)$, which in order for the $O(x^4)$ terms to vanish places a restriction on the matrix $\mathbf{A}$. Splitting the symmetric and antisymmetric parts of $\mathbf{A}$ into condensed-condensed, condensed-uncondensed, uncondensed-condensed, and uncondensed-uncondensed submatrices

$$\mathbf{A}=\begin{bmatrix}A^s_{cc} & A^s_{cu}\\ A^s_{uc} & A^s_{uu}\end{bmatrix}+\begin{bmatrix}A^a_{cc} & A^a_{cu}\\ A^a_{uc} & A^a_{uu}\end{bmatrix}$$

showed the restrictions on $\mathbf{A}$ to be $\sum_{\nu>c}A^{s,a}_{\mu\nu}z^{\nu}\sim O(1)$ and $\sum_{\nu\le c}A^{s,a}_{\mu\nu},m^{\nu}\sim O(1/\sqrt{N})$. Assuming the matrix $\mathbf{q}^{-1}\otimes\mathbb{1}+(2i/\alpha)\mathbf{R}\otimes\mathbf{A}^s_{uu}$ to be positive definite, the result of the Gaussian integral has the form e^{Φ} where

$$\Phi=\ln\left[\frac{\exp\left[-\frac{1}{2}\boldsymbol{\Upsilon}\cdot\left[\mathbf{q}^{-1}\otimes\mathbb{1}+\frac{2i}{\alpha}\mathbf{R}\otimes\mathbf{A}^s_{uu}\right]^{-1}\boldsymbol{\Upsilon}\right]}{\left[\det(\mathbf{q}\otimes\mathbb{1})\det\left[\mathbf{q}^{-1}\otimes\mathbb{1}+\frac{2i}{\alpha}\mathbf{R}\otimes\mathbf{A}^s_{uu}\right]\right]^{1/2}}\right] \tag{19}$$

$$\times\sum_{\mu>s}\ln\left[1-\frac{1}{2}\frac{\int d\mathbf{z}\exp\left[-\frac{1}{2}\mathbf{z}\cdot\left[\mathbf{q}^{-1}\otimes\mathbb{1}+\frac{2i}{\alpha}\mathbf{R}\otimes\mathbf{A}^s_{uu}\right]\mathbf{z}-i\boldsymbol{\Upsilon}\cdot z\right]\Gamma^2_{\mu}(x,y,\mathbf{m},R,\mathbf{z})}{\int d\mathbf{z}\exp\left[-\frac{1}{2}\mathbf{z}\cdot\left[\mathbf{q}^{-1}\otimes\mathbb{1}+\frac{2i}{\alpha}\mathbf{R}\otimes\mathbf{A}^s_{uu}\right]\mathbf{z}-i\boldsymbol{\Upsilon}\cdot\mathbf{z}\right]}\right],$$

$$\Upsilon^{\nu}_{a}=\sum_{\mu\le c}\zeta^{\mu}(xA^s_{\mu\nu}+yA^a_{\mu\nu})\frac{\delta_{\alpha 1}}{\sqrt{N}}+\frac{2}{\alpha}\sqrt{N}R^{\alpha}m^{\mu}A^s_{\mu\nu}\ , \tag{20}$$

$$\Gamma_{\mu}(x,y,\mathbf{m},R,\mathbf{z})=\left[\sum_{\nu>c}(xA^s_{\mu\nu}+yA^a_{\mu\nu})\frac{z^{\nu}_1}{\sqrt{N}}+\sum_{\nu\le c}(xA^s_{\mu\nu}+yA^a_{\mu\nu})m^{\nu}\right]+\frac{2}{\alpha}\sum_{\alpha}R^{\alpha}\left[\sum_{\nu>c}A^s_{\mu\nu}\frac{s^{\alpha}_i z^{\nu}_{\alpha}}{\sqrt{N}}+\sum_{\nu\le c}A^s_{\mu\nu}s^{\alpha}_i m^{\nu}\right]\ . \tag{21}$$

We split Φ into extensive and the intensive parts; $\Phi=N\Omega-\mathcal{R}$:

$$\Omega=-\frac{2}{\alpha^2}\sum_{\alpha,\beta}\sum_{\mu\rho\le c}\sum_{\nu\eta>c}R^{\alpha}m^{\mu}A^{s}_{\mu\nu}(\Lambda^{-1})^{\alpha\beta}_{\nu\eta}A^{s}_{\eta\rho}m^{\rho}R^{\beta}-\frac{1}{2N}\det(\mathbf{q}\otimes\mathbb{1})\det(\Lambda)\ , \tag{22}$$

$$\begin{aligned}\mathcal{R}=&\frac{1}{\alpha}\sum_{\beta}\sum_{\mu,\rho\le c}\sum_{\nu,\eta>c}[\xi^{\mu}(xA^{s}_{\mu\nu}+yA^{a}_{\mu\nu})](\Lambda^{-1})^{1\beta}_{\nu\eta}A^{s}_{\eta\rho}m^{\rho}R^{\beta}\\&+\frac{1}{\alpha}\sum_{\alpha}\sum_{\mu,\rho\le c}\sum_{\nu,\eta>c}R^{\alpha}m^{\mu}A^{s}_{\mu\nu}(\Lambda^{-1})^{\alpha 1}_{\nu\eta}[\xi^{\rho}(xA^{s}_{\rho\eta}+yA^{a}_{\rho\eta})]^{\dagger}\\&+\frac{1}{2N}\sum_{\mu,\rho\le c}\sum_{\nu,\eta>c}[\xi^{\mu}(xA^{s}_{\mu\nu}+yA^{a}_{\mu\nu})](\Lambda^{-1})^{11}_{\nu\eta}[\xi^{\rho}(xA^{s}_{\rho\eta}+yA^{a}_{\rho\eta})]^{\dagger}\\&+\tfrac{1}{2}\sum_{\mu>c}\frac{\int d\mathbf{z}\exp\left[-\frac{1}{2}\mathbf{z}\left[\mathbf{q}^{-1}\otimes\mathbb{1}+\frac{2i}{\alpha}\mathbf{R}\otimes\mathbf{A}_{uu}\right]\mathbf{z}-i\boldsymbol{\Upsilon}\mathbf{z}\right]\Gamma^{2}_{\mu}(x,y,\mathbf{m},R,\mathbf{z})}{\int d\mathbf{z}\exp\left[-\frac{1}{2}\mathbf{z}\left[\mathbf{q}^{-1}\otimes\mathbb{1}+\frac{2i}{\alpha}\mathbf{R}\otimes\mathbf{A}_{uu}\right]\mathbf{z}-i\boldsymbol{\Upsilon}\mathbf{z}\right]}\ ,\end{aligned} \tag{23}$$

$$\Lambda^{\alpha\beta}_{\mu\nu}=q^{-1}_{\alpha\beta}\delta_{\mu\nu}+\frac{2i}{\alpha}R^{\alpha}\delta_{\alpha\beta}A^{s}_{\mu\nu}\ . \tag{24}$$

Notice that since Λ is symmetric in the indices μ,ν, the antisymmetric parts in the first two lines of the expression for $\mathcal{R}$ cancel, and symmetric and antisymmetric parts in the third line decouple.

In order to facilitate the spin average, we now introduce the following representation of unity:

$$1=\int dq_{\alpha\beta}\delta[q_{\alpha\beta}-q_{\alpha\beta}(\mathbf{s})]=\left[\frac{N}{2\pi}\right]^{n^2}\int dq_{\alpha\beta}d\hat{q}_{\alpha\beta}\exp\{iN\hat{q}_{\alpha\beta}[q_{\alpha\beta}-q_{\alpha\beta}(\mathbf{s})]\} \tag{25}$$

Hence the noise distributions (15) become

$$\mathcal{D}_{\xi}[z^{s},z^{a}]\sim\lim_{n\to 0}\int d\mathbf{R}\,d\mathbf{M}\,d\mathbf{q}\,d\hat{\mathbf{q}}\exp[N\Psi(\mathbf{R},\mathbf{M},\mathbf{q},\hat{\mathbf{q}})]\int dx\,dy\,\exp(ixz^{s}+iyz^{a})\left\langle\exp\left[-\mathcal{R}(x,y)-i\sum_{\alpha}\sum_{\mu\le c}\xi^{\mu}M^{a}_{\mu}s^{\alpha}_{i}\right]\right\rangle_{\{s^{\alpha}\}}\ , \tag{26}$$

where

$$\Psi=i\sum_{\alpha}(\mathbf{m}\mathbf{M}^{\alpha}+rR^{\alpha})+i\sum_{\alpha,\beta}q_{\alpha\beta}\hat{q}_{\alpha\beta}+\Omega(\mathbf{R},\mathbf{q})+\frac{1}{N}\sum_{k\neq i}\ln\left\langle\left\langle\exp\left[-i\left[\sum_{\alpha}\sum_{\mu\le c}\xi^{\mu}_{k}M^{\alpha}_{\mu}s^{\alpha}_{i}+\sum_{\alpha,\beta}\hat{q}_{\alpha\beta}s^{\alpha}_{k}s^{\beta}_{k}\right]\right]\right\rangle_{\xi}\right\rangle_{\{s^{\alpha}\}}\ . \tag{27}$$

The integrals over $\mathbf{R},\mathbf{M},\mathbf{q},\hat{\mathbf{q}}$ are to be performed by the saddle point method. The saddle point relations in turn allow us to evaluate the (intensive) integrals over x and y determining the noise distribution.

B. Replica-symmetric saddle points

We assume that the relevant saddle points are replica symmetric, i.e., that

$$q_{\alpha\beta}=\delta_{\alpha\beta}+q(1-\delta_{\alpha\beta})\qquad\forall\alpha,\beta\ ,$$

$$\hat{q}_{\alpha\beta}=\hat{q}(1-\delta_{\alpha\beta})\quad\forall\alpha,\beta\ ,$$

$$R^{\alpha}=R\,\forall\alpha\ ,$$

and

$$\mathbf{M}^{\alpha}=\mathbf{M}\ \forall\alpha$$

(N.B. the diagonal elements $\hat{q}_{\alpha\alpha}$ vanish automatically). At this stage it is convenient to make the change of variables $\rho=-(2i/\alpha)R$, $\hat{q}=(i/2)\lambda^2$, and $M=i\mu$. The matrix $\delta_{\alpha\beta}+q(1-\delta_{\alpha\beta})$ has one eigenvalue $1+q(n-1)$ and $n-1$ eigenvalues $q-q$; hence we obtain

$$\ln\det(\mathbf{q}\otimes\mathbb{1})+\ln\det(\Lambda)=\mathrm{Tr}\ln[\delta_{\mu\nu}-\rho A^{s}_{\mu\nu}(1-q+nq)]+(n-1)\mathrm{Tr}\ln[\delta_{\mu\nu}-\rho A^{s}_{\mu\nu}(1-q)] \tag{28}$$

and

$$\Lambda^{-1\alpha\beta}_{\mu\nu}=\delta_{\alpha\beta}(1-q)[\delta_{\mu\nu}-\rho A^{s}_{\mu\nu}(1-q)]^{-1}+q[\delta_{\mu\lambda}-\rho A_{\mu\lambda}s(1-q)]^{-1}[\delta_{\lambda\nu}-\rho A^{s}_{\lambda\nu}(1-q+nq)]^{-1} . \tag{29}$$

The last term in Ψ (27) becomes

$$\left\langle \ln \int Dw \exp\left[n \ln\cosh\left[\sum_{\nu\le c} \xi^{\nu}\mu^{\nu}+\lambda w \right]\right]\right\rangle_{\xi} , \quad Dw=\exp\left[-\frac{1}{2}w^2\right]\frac{dw}{\sqrt{2\pi}} \tag{30}$$

The saddle point equations $\partial\Psi/\partial M^{\alpha}_{\mu}=0$, $\partial\Psi/\partial\hat{q}_{\alpha\beta}=0$, $\partial\Psi/\partial R^{\alpha}=0$, $\partial\Psi/\partial q_{\alpha\beta}=0$ give the following:

$$m^{\mu}=\left\langle \frac{\int Dw\, \xi^{\mu}\tanh\left[\sum_{\nu\le c}\xi^{\nu}\mu^{\nu}+\lambda w\right]\cosh^{n}\left[\sum_{\nu\le c}\xi^{\nu}\mu^{\nu}+\lambda w\right]}{\int Dw \cosh^{n}\left[\sum_{\nu\le c}\xi^{\nu}\mu^{\nu}+\lambda w\right]}\right\rangle_{\xi} , \tag{31}$$

$$q=\left\langle \frac{\int Dw\, \tanh^2\left[\sum_{\nu\le c}\xi^{\nu}\mu^{\nu}+\lambda w\right]\cosh^{n}\left[\sum_{\nu\le c}\xi^{\nu}\mu^{\nu}+\lambda w\right]}{\int Dw \cosh^{n}\left[\sum_{\nu\le c}\xi^{\nu}\mu^{\nu}+\lambda w\right]}\right\rangle_{\xi} , \tag{32}$$

$$\begin{aligned} r=&\frac{1}{p}\mathrm{Tr}\{\mathbf{A}^{s}_{uu}[\mathbb{1}-\rho(1-q)(1-q+nq)\mathbf{A}^{s}_{uu}][\mathbb{1}-\rho(1-q+nq)\mathbf{A}^{s}_{uu}]^{-1}[\mathbb{1}-\rho(1-q)\mathbf{A}^{s}_{uu}]^{-1}\} \\ &+\frac{1}{\alpha n}\frac{\partial}{\partial\rho}\rho^2\mathbf{m}\cdot\mathbf{A}^{s}_{cu}\mathbf{\Lambda}^{-1}(q,\rho)\mathbf{A}^{s}_{uc}\mathbf{m} , \end{aligned} \tag{33}$$

$$\lambda^2=\frac{\alpha q\rho^2}{p}\mathrm{Tr}\{\mathbf{A}^{s}_{uu}\mathbf{A}^{s}_{uu}[\mathbb{1}-\rho(1-q+nq)\mathbf{A}^{s}_{uu}]^{-1}[\mathbb{1}-\rho(1-q)\mathbf{A}^{s}_{uu}]^{-1}\}+\frac{1}{n(n-1)}\rho^2\frac{\partial}{\partial q}\mathbf{m}\cdot\mathbf{A}^{s}_{cu}\mathbf{\Lambda}^{-1}(q,\rho)\mathbf{A}^{s}_{uc}\mathbf{m} , \tag{34}$$

and the extensive exponent Ψ, evaluated at the saddle point, is given by

$$\begin{aligned} \frac{1}{n}\Psi(\boldsymbol{\mu},\rho,q,\lambda)=&-\mathbf{m}\cdot\boldsymbol{\mu}-\tfrac{1}{2}\alpha\rho-\tfrac{1}{2}(n-1)\lambda^2 q-\frac{1}{2nN}\mathrm{Tr}\{\ln[\mathbb{1}-\rho\mathbf{A}^{s}_{uu}(1-q+nq)]\} \\ &-\frac{n-1}{2nN}\mathrm{Tr}\{\ln[\mathbb{1}-\rho\mathbf{A}^{s}_{uu}(1-q)]\}+\frac{\rho^2}{n}\mathbf{m}\cdot\mathbf{A}^{s}\sum_{\alpha,\beta}\mathbf{\Lambda}^{-1\alpha\beta}\mathbf{A}^{s}\mathbf{m} \\ &+\frac{1}{n}\left\langle \ln\int Dw\exp\left[n\ln\cosh\left[\sum_{\mu\le c}\xi^{\mu}\mu^{\mu}+\lambda w\right]\right]\right\rangle_{\xi} . \end{aligned} \tag{35}$$

$\mathbf{\Lambda}^{-1}$ has the general form $\mathbf{B}\delta_{\alpha\beta}+\mathbf{C}$. Such matrices have the properties

$$\lim_{n\to 0}\frac{1}{n}\sum_{\alpha,\beta}[\mathbf{B}\delta_{\alpha\beta}+\mathbf{C}]=\mathbf{B} , \quad \lim_{n\to 0}\frac{1}{n(n-1)}\sum_{\alpha,\beta}[\mathbf{B}\delta_{\alpha\beta}+\mathbf{C}]=-\mathbf{B} . \tag{36}$$

Therefore in the limit $n\to 0$ our saddle point equations become

$$m^{\mu}=\left\langle \int Dw\, \xi^{\mu}\tanh\left[\sum_{\nu\le c}\xi^{\nu}\mu^{\nu}+\lambda w\right]\right\rangle_{\xi} , \tag{37}$$

$$q=\left\langle \int Dw\, \tanh^2\left[\sum_{\nu\le c}\xi^{\nu}\mu^{\nu}+\lambda w\right]\right\rangle_{\xi} , \tag{38}$$

$$\begin{aligned} r=&\frac{1}{p}\mathrm{Tr}\{\mathbf{A}^{s}_{uu}[\mathbb{1}-\rho(1-q)^2\mathbf{A}^{s}_{uu}][\mathbb{1}-\rho(1-q)\mathbf{A}^{s}_{uu}]^{-1}[\mathbb{1}-\rho(1-q)\mathbf{A}^{s}_{uu}]^{-1}\} \\ &+\frac{1-q}{\alpha}\frac{\partial}{\partial\rho}\rho^2\mathbf{m}\cdot\mathbf{A}^{s}_{cu}[\mathbb{1}-\rho\mathbf{A}^{s}_{uu}(1-q)]^{-1}\mathbf{A}^{s}_{uc}\mathbf{m} , \end{aligned} \tag{39}$$

$$\lambda^2=\frac{\alpha q\rho^2}{p}\mathrm{Tr}\{\mathbf{A}^{s}_{uu}\mathbf{A}^{s}_{uu}[\mathbb{1}-\rho(1-q)\mathbf{A}^{s}_{uu}]^{-1}[\mathbb{1}-\rho(1-q)\mathbf{A}^{s}_{uu}]^{-1}\}+\rho^2\mathbf{m}\cdot\mathbf{A}^{s}_{cu}[\mathbb{1}-\rho(1-q)\mathbf{A}^{s}_{uu}]^{-2}\mathbf{A}^{2}_{uc}\mathbf{m} . \tag{40}$$

These equations are to be solved for $\mathbf{\Lambda}$ positive definite, which is necessary for the Gaussian integral to be well defined, possibly placing restrictions on the eigenvalues of $\mathbf{A}$.

C. Shape of the intrinsic noise distribution

The shape of the noise distribution is now determined by the intensive terms in the integrand, with the values of μ,q,ρ,λ evaluated in the saddle point. We leave the details of the calculation of $\mathcal{R}$ to Appendix A; the result is

$$\mathcal{D}_\xi[z^s,z^a]\sim\int dx\,dy\exp(ixz^s+iyz^a)\left\langle\exp\left[-\mathcal{R}(x,y)+\sum_{\nu\le c}\xi^\nu\mu^\nu\sum_\alpha s_i^\alpha\right]\right\rangle_{\{s^\alpha\}} \tag{41}$$

$$\sim\int\frac{dx}{2\pi}\frac{dy}{2\pi}\exp(ixz^s+iyz^a)\exp(-\tfrac{1}{2}y^2\mathcal{G})$$
$$\times\left\langle\exp\left[\sum_{\nu\le c}\xi^\nu\mu^\nu\sum_\alpha s_i^\alpha-ix\mathcal{B}-ix\rho\left[\alpha\tilde{r}s_i^1+\frac{\tilde{\lambda}^2}{\rho^2}\sum_{\alpha>1}s_i^\alpha\right]-x^2\mathcal{E}-i\rho x\sum_\alpha\mathcal{F}_\alpha s_i^\alpha\right]\int Dw\exp\left[w\Gamma\sum_\alpha s_i^\alpha\right]\right\rangle_{\{s^\alpha\}}, \tag{42}$$

where

$$\mathcal{G}=-\alpha\tilde{p}-2\rho\mathbf{m}\cdot\mathbf{A}_{cu}^a\,\mathbf{A}_{uu}^a\,\mathbf{A}_{uu}^{s\;-1}[\tilde{\mathcal{R}}+(n-1)\mathcal{A}]\mathbf{m}-\rho^2\mathbf{m}\cdot\mathbf{A}_{cu}^s\tilde{\mathcal{P}}\mathbf{A}_{uc}^s\mathbf{m}-\mathbf{m}\cdot\mathbf{A}_{cu}^a\,\mathbf{A}_{uc}^a\mathbf{m}\;, \tag{43}$$

$$\mathcal{B}=\rho\boldsymbol{\xi}\cdot[\tilde{\mathcal{R}}_{cc}+(n-1)\mathcal{A}_{cc}]\mathbf{m}\;, \tag{44}$$

$$\mathcal{E}=\frac{1}{2}\mathbf{m}\cdot\{\mathbf{A}_{cu}^s+\rho[\tilde{\mathcal{R}}_{cu}+(n-1)\mathcal{A}_{cu}]\}\{\mathbf{A}_{uc}^s+\rho[\tilde{\mathcal{R}}_{uv}+(n-1)\mathcal{A}_{uc}]\}\mathbf{m}+\frac{\alpha\tilde{r}}{2}\;, \tag{45}$$

$$\mathcal{F}=\mathbf{m}\cdot\{\mathbf{A}_{cu}^s+\rho[\tilde{\mathcal{R}}_{cu}+(n-1)\mathcal{A}_{cu}]\}\{\mathbf{A}_{uc}^s+\rho[\tilde{\mathcal{R}}_{uc}+(n-1)\mathcal{A}_{uc}]\}\mathbf{m}\;, \tag{46}$$

$$\Gamma^2=\lambda^2=\tilde{\lambda}^2+\rho^2\mathbf{m}\cdot\{\mathbf{A}_{cu}^s+\rho[\tilde{\mathcal{R}}_{cu}+(n-1)\mathcal{A}_{cu}]\}\{\mathbf{A}_{uc}^s+\rho[\tilde{\mathcal{R}}_{uc}+(n-1)\mathcal{A}_{uc}]\}\mathbf{m}\;, \tag{47}$$

$$\frac{\tilde{\lambda}^2}{\rho^2}=\frac{1}{N}\sum_{\mu,\nu,\lambda>c}A_{\lambda\mu}^s(\mathbf{\Lambda}^{-1})_{\mu\nu}^{\alpha\ne\beta}A_{\nu\lambda}^s=\frac{1}{N}\mathrm{Tr}\mathcal{A}$$
$$=\frac{\alpha q}{p}\mathrm{Tr}\{\mathbf{A}_{uu}^s\,\mathbf{A}_{uu}^s[\mathbb{1}-\rho(1-q+nq)\mathbf{A}_{uu}^s]^{-1}[\mathbb{1}-\rho(1-q)\mathbf{A}_{uu}^s]^{-1}\}\;, \tag{48}$$

$$\alpha\tilde{r}=\frac{1}{N}\sum_{\mu,\nu,\lambda>c}A_{\lambda\mu}^s(\mathbf{\Lambda}^{-1})_{\mu\nu}^{\alpha=\beta}A_{\mu\lambda}^s=\frac{1}{N}\mathrm{Tr}\tilde{\mathcal{R}} \tag{49}$$
$$=\frac{1}{p}\mathrm{Tr}\{\mathbf{A}_{uu}^s\,\mathbf{A}_{uu}^s[\mathbb{1}-\rho(1-q)(1-q+nq)\mathbf{A}_{uu}^s][\mathbb{1}-\rho(1-q+nq)\mathbf{A}_{uu}^s]^{-1}[\mathbb{1}-\rho(1-q)\mathbf{A}_{uu}^s]^{-1}\}\;,$$

$$\alpha\tilde{p}=\frac{1}{N}\sum_{\mu,\nu,\lambda>c}A_{\lambda\mu}^a(\mathbf{\Lambda}^{-1})_{\mu\nu}^{11}A_{\nu\lambda}^a=\frac{1}{N}\mathrm{Tr}\tilde{\mathcal{P}}\;. \tag{50}$$

Carrying out the spin average gives

$$\mathcal{D}_\xi[z^s,z^a]=\lim_{n\to0}\int\frac{dx}{2\pi}\frac{dy}{2\pi}\exp(ixz^s+iyz^a)\exp(-\tfrac{1}{2}y^2\mathcal{G})\exp(-x^2\mathcal{E}-ix\mathcal{B})\int Dw$$
$$\times\cosh^{n-1}\left[\sum_{\nu\le c}\xi^\nu\mu^\nu-ix\frac{\tilde{\lambda}^2}{\rho}-i\rho x\mathcal{F}+w\Gamma\right]\cosh\left[\sum_{\nu\le c}\xi^\nu\mu^\nu-ix\rho\alpha\tilde{r}-i\rho x\mathcal{F}+w\Gamma\right]. \tag{51}$$

We can carry out the integral over y immediately (assuming $\mathcal{G}$ to be positive); however, the $\cosh^{n-1}(\cdots)$ will cause a divergence for $n\to0$. However, if we define

$$\Delta=\alpha\rho\tilde{r}-\frac{\tilde{\lambda}^2}{\rho}=\frac{\alpha\rho(1-q)}{p}\mathrm{Tr}\{\mathbf{A}_{uu}^s[\mathbb{1}-\rho(1-q)\mathbf{A}_{uu}^s]^{-1}\mathbf{A}_{uu}^s\}\;, \tag{52}$$

the cosh terms in (52) become

$$\cos(\Delta x)\cosh\left[\sum_{\nu\le c}\xi^\nu\mu^\nu-ix\frac{\tilde{\lambda}^2}{\rho}-i\rho x\mathcal{F}+w\Gamma\right]-i\sin(\Delta x)\sinh\left[\sum_{\nu\le c}\xi^\nu\mu^\nu-ix\frac{\tilde{\lambda}^2}{\rho}-i\rho x\mathcal{F}+w\Gamma\right], \tag{53}$$

so that the possible divergence eliminated by the contour shift $\hat{y}'=w-(ix/\Gamma)(\tilde{\lambda}^2/\rho+\rho\mathcal{F})$, giving

$$\mathcal{D}_\xi[z^s,z^a]=(2\pi\mathcal{G})^{-1/2}\exp(-\tfrac{1}{2}z^a\mathcal{G}^{-1}z^a)\int\frac{dx}{2\pi}\exp\left[ix(z^s-\mathcal{B})-\frac{\Delta}{2\rho}x^2\right]\int D\hat{y}'\exp\left[-ix\hat{y}'\frac{\Gamma}{\rho}\right]$$
$$\times\frac{1}{2}\left[(e^{i\Delta x}+e^{-ix\Delta})-(e^{i\Delta x}-e^{-ix\Delta})\tanh\left[\sum_{\nu\le c}\xi^\nu\mu^\nu+\hat{y}'\Gamma\right]\right], \tag{54}$$

where we have used the relations $\tilde{\lambda}^2/\rho+\rho\mathcal{F}=\Gamma^2/\rho$ and $\mathcal{E}-(1/2)(\tilde{\lambda}^2/\rho+\rho\mathcal{F})=\Delta/2\rho$. The remaining integral over x is easily performed using $1+\Gamma^2/\rho\Delta=2\mathcal{E}\rho/\Delta$ and $-1/2\mathcal{E}+\rho/\Delta=\Gamma^2/2\rho\mathcal{E}\Delta$ and (again) a rescaling $y=\sqrt{2\mathcal{E}\rho/\Delta}[\hat{y}'-(\Gamma/2\mathcal{E}\rho)(z^s\pm\Delta-\mathcal{B})]$, giving the final result

$$\mathcal{D}_{\xi}[z^s,z^a]=\frac{e^{-(1/2)z^a g^{-1}z^a}}{4\pi\sqrt{2\mathcal{G}\mathcal{E}}}\left\{\exp\left[-\frac{(z^s-\mathcal{B}+\Delta)^2}{4\mathcal{E}}\right]\left\{1-\int Dy\tanh\left[\sum_{\nu\le c}\xi^\nu\mu^\nu+y\Gamma\left[\frac{\Delta}{2\mathcal{E}\rho}\right]^{1/2}+(z^s+\Delta-\mathcal{B})\frac{\Gamma^2}{2\mathcal{E}\rho}\right]\right\}\right.$$
$$\left.+\exp\left[-\frac{(z^s-\mathcal{B}-\Delta)^2}{4\mathcal{E}}\right]\left\{1+\int Dy\tanh\left[\sum_{\nu\le c}\xi^\nu\mu^\nu+y\Gamma\left[\frac{\Delta}{2\mathcal{E}\rho}\right]^{1/2}+(z^s-\Delta-\mathcal{B})\frac{\Gamma^2}{2\mathcal{E}\rho}\right]\right\}\right\}. \tag{55}$$

Figures 1 and 2 show the noise distributions as a function of z^s for four values of z^a and

$$\xi=\begin{pmatrix}1\\1\end{pmatrix}$$

as predicted by our theory (solid line) and as measured using spin simulations (histograms). The simulations were carried out for a system of 40 000 spins (neurons), using the block diagonal matrix (76) (see Sec. IV). Figure 1 corresponds to the values of parameters $a=1$ (giving rise to cyclical behavior), $\alpha=0.0256$, $m_1=-0.2173$, $m_2=-0.5511$, and $r=3.700\,631$, while Fig. 2 corresponds to the parameter values $a=2$ (converging towards the fixed point $\mathbf{m}=0, r>1$), $\alpha=0.0512$, $m_1=-0.007\,35$, $m_2=0.004\,35$, and $r=4.173\,877$.

From Figs. 1 and 2 we can see that in the first case when cyclical behavior is observed our model appears to underestimate the distribution at extreme values of z^a while overestimating for z^a around zero. This could be partially caused by the finite size of the bins of the histogram in the z^a direction, and by other finite size effects. [The number of spins needed to obtain good statistics given that one must specify (i) a value of ξ, (ii) an interval of z^s, and (iii) an interval of z^a, is unfeasibly large.]

Both sets of distributions "look" approximately Gaussian. In fact the results of numerous simulations have revealed that it is extremely difficult to force a network with nonsymmetric interactions to iterate towards a state where the noise distribution is strongly non-Gaussian. In contrast, the symmetric case [3] nonretrieval states $(\mathbf{m}=\mathbf{0})$ have double peaked noise distributions. Al-

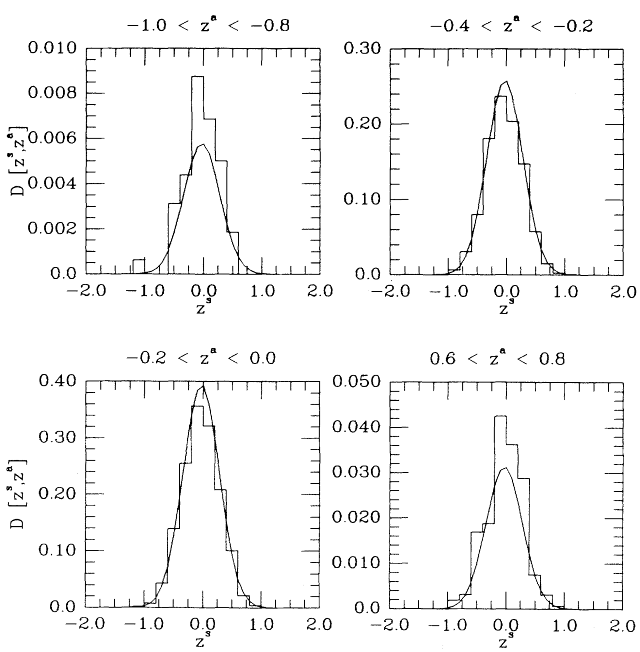

FIG. 1. Plots showing sections through the noise distribution at values of z^a near $z^a=-0.9$, $z^a=-0.3$, $z^a=0.01$, and $z^a=0.7$, for a model with block diagonal $\mathbf{A}$ (76) (see Sec. IV) with $a=1$ (giving rise to cyclical behavior), $\alpha=0.0256$, $m_1=-0.2173$, $m_2=-0.5511$, and $r=3.700\,631$. The solid lines are theoretical predictions, while the histograms are taken from simulations by counting the number of sites with z^s, z^a in a range ± 0.1 from the center.

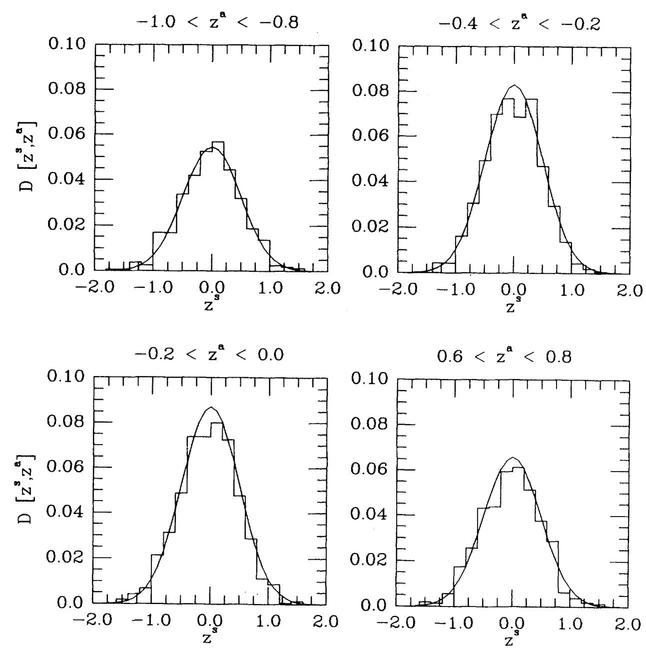

FIG. 2. Plots showing sections through the noise distribution at values of z^a near $z^a=-0.9$, $z^a=-0.3$, $z^a=-0.1$, and $z^a=0.7$, for a model with block diagonal $\mathbf{A}$ (76) with $a=2$ (see Sec. IV) (converging towards the fixed point $\mathbf{m}=0$, $r>1$), $\alpha=0.0512$, $m_1=-0.007\,35$, $m_2=0.004\,35$, and $r=4.173\,877$. The solid lines are theoretical predictions, while the histograms are taken from simulations by counting the number of sites with z^s, z^a in a range ± 0.1 from the center.

though the theory does not describe the shape of the distributions exactly in these cases [16], it shows a good qualitative fit to the clearly non-Gaussian shape, in contrast to the time dependent Ginzburg-Landau approaches [10,11]. The lack of non-Gaussian shaped noise distributions networks with nonsymmetric interactions implies that symmetry plays an important role in building up the correlations between the system state and the uncondensed patterns at the microscopic level, or equivalently, that the macroscopic dynamics does not lead the system to a region of $(\mathbf{m},r)$ space where the noise distribution is strongly non-Gaussian $(q\to 1)$.

D. Special cases and critical surfaces

The noise distribution (55) in general can only be calculated numerically. For specific choices of $\mathbf{A}$, however, it simplifies considerably. In the absence of condensed-uncondensed couplings, for instance, $\mathcal{G}=-\alpha\tilde{p}$, $\mathcal{B}=0$, $\mathcal{E}=\alpha\tilde{r}/2$, $\mathcal{F}=0$, and $\Gamma^2=\tilde{\lambda}^2$. In the absence of antisymmetric components to $\mathbf{A}$ we find that $\mathcal{G}=0$, and the Gaussian function of z^a becomes $\delta(z^a)$. For $\mathbf{A}=\mathbb{1}$ (the Hopfield model) the noise distribution reduces to the result given in [5].

There are also several regions [in the space of $(\mathbf{m},r)$] where the noise distribution takes on a special form, or where significant transitions occur.

1. Gaussian noise

In previous papers the noise distribution has often been assumed to have a Gaussian shape [11]. Here we can see that this requires $\Delta=0$, which implies that

$$q\delta_{\mu\nu}=\delta_{\mu\nu}-\rho(1-q)^2A^s_{\mu\nu}\ , \tag{56}$$

requiring $A^s_{\mu\nu}=A\delta_{\mu\nu}$ and $\rho=1/A(q-1)$. Tracing this back through the saddle point equations leads to the requirement $r=1$, which explains why the Gaussian noise approximation is only reasonable in the region of $(\mathbf{m},r)$ space where retrieval occurs.

2. $q=0, m=0$

From the saddle point equations we can see that $m=0$ implies $\mu=0$. Expanding the equation for q around $q=0$ therefore gives

$$q\simeq\lambda^2\int Dw\,w^2\simeq\frac{\alpha q\rho^2}{p}\mathrm{Tr}[\,\mathbf{A}^s_{uu}\,\mathbf{A}^s_{uu}(\mathbb{1}-\rho\,\mathbf{A}^s_{uu})^{-2}]\ . \tag{57}$$

The only solution is $q=0$ if $1>(\alpha\rho^2/p)\mathrm{Tr}[\,\mathbf{A}^s_{uu}\,\mathbf{A}^s_{uu}(\mathbb{1}-\rho\,\mathbf{A}^s_{uu})^{-2}]$. The saddle point equation for r at $q=0$: $r=r_c=1/p\mathrm{Tr}[\,\mathbf{A}^s_{uu}(\mathbb{1}-\rho\,\mathbf{A}^s_{uu})^{-1}]$ subsequently defines the value of r below which $q=\mathbf{m}=0$ is the only solution. In this region of phase space ($\mathbf{m}=0$, $r<r_c$) the noise distribution is a sum of two Gaussians:

$$\mathcal{D}_\xi[z^s,z^a]=\frac{\exp\left[\dfrac{(z^a)^2}{2\alpha\tilde{p}}\right]}{4\pi\alpha\sqrt{-\tilde{r}\tilde{p}}}\left[\exp\left[-\frac{(z^s+\alpha\rho\tilde{r})^2}{2\alpha\tilde{r}}\right]+\exp\left[-\frac{(z^s-\alpha\rho\tilde{r})^2}{2\alpha\tilde{r}}\right]\right]\ . \tag{58}$$

3. Freezing line

The freezing line occurs when the number of microscopic states contributing to the macroscopic state goes from an extensive large number to an extensively small number (equivalent to a transition to a macroscopic state with negative entropy):

$$\lim_{N\to\infty}\frac{1}{N}\ln\sum_{s}\delta(m-m(\mathbf{s}))\delta(r-r(\mathbf{s}))=0\ . \tag{59}$$

This is the first sign that something has gone wrong with the replica method. We can relate the freezing line to the saddle point problem encountered in the calculation of the intrinsic noise distribution. In order to average over the patterns ξ we use $\ln Z=\lim_{n\to 0}(1/n)(Z^n-1)$ to cast (59) into the familiar form

$$\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{nN}\left[\left\langle\left\langle\prod_{a=1}^{n}\delta(m-m(\mathbf{s}^\alpha))\delta(r-r(\mathbf{s}^\alpha))\right\rangle\right\rangle-1\right]=\lim_{n\to 0}\frac{1}{n}\Psi=-\ln 2\ . \tag{60}$$

Using (27) this leads in the replica-symmetric (RS) approximation to

$$\begin{aligned}-\ln 2=&-\mathbf{m}\cdot\boldsymbol{\mu}-\frac{1}{2}\alpha\rho+\frac{1}{2}\lambda^2q+\frac{\rho q}{2N}\mathrm{Tr}\{\,\mathbf{A}^s_{uu}[\mathbb{1}-\rho\,\mathbf{A}^s_{uu}(1-q)]^{-1}\}\\&-\frac{1}{2N}\mathrm{Tr}\{\ln[\mathbb{1}-\rho\,\mathbf{A}^s_{uu}(1-q)]\}+\rho^2(1-q)\mathbf{m}\cdot\mathbf{A}^s[\mathbb{1}-\rho(1-q)\,\mathbf{A}^s_{uu}]^{-1}\,\mathbf{A}^s_{uc}\mathbf{m}\\&+\left\langle\int Dy\,\ln\cosh\left[\sum_{\mu\le c}\xi^\mu\mu^\mu+\lambda y\right]\right\rangle_\xi\ .\end{aligned} \tag{61}$$

This freezing line defined by (61) occurs close to, but not exactly on, the $q=1$ line. The above equation is to be solved numerically along with the saddle point equations.

4. $q=1$

In the $q\simeq 1$ region of $(\mathbf{m},r)$ space we can use the asymptotic form of the saddle point equations to derive a simpler expression for the noise distribution. We expand the saddle point equations in $\epsilon=1-q$ in order to determine where the $q=1$ line occurs. We first note that $\lambda\propto\rho$. From the saddle point equation relating r and ρ we note that $\rho\propto\epsilon^{-1}$ and hence $\lambda\propto\epsilon^{-1}$. Therefore $\tanh(\lambda\hat{y}+\boldsymbol{\xi}\cdot\boldsymbol{\mu})=\mathrm{sgn}(\lambda\hat{y}+\boldsymbol{\xi}\cdot\boldsymbol{\mu})$, which we use to simplify our saddle point equations for $\mathbf{m}$ and q:

$$m^{\mu}=\left\langle \xi^{\mu}\mathrm{erf}\left[\frac{\boldsymbol{\xi}\cdot\boldsymbol{\mu}}{\sqrt{2\lambda}}\right]\right\rangle_{\xi}, \tag{62}$$

$$\begin{aligned}-\epsilon=q-1&=\left\langle \int D\hat{y}\,\tanh^2(\lambda\hat{y}+\boldsymbol{\xi}\cdot\boldsymbol{\mu})-1\right\rangle_{\xi}\\
&=\left\langle\left[\frac{e^{-(1/2)y^2}}{\sqrt{2\pi}}\frac{1}{\lambda}\tanh(\lambda\hat{y}+\boldsymbol{\xi}\cdot\boldsymbol{\mu})\right]\right\rangle_{\xi}-\frac{2}{\lambda}\left\langle\int\frac{dy}{\sqrt{2\pi}}ye^{-(1/2)y^2}\tanh(\lambda\hat{y}+\boldsymbol{\xi}\cdot\boldsymbol{\mu})\right\rangle_{\xi}\\
&=-\frac{2}{\lambda}\left\langle\int_{-\boldsymbol{\xi}\cdot\boldsymbol{\mu}/\lambda}^{\infty}\hat{y}\,D\hat{y}-\int_{-\infty}^{-\boldsymbol{\xi}\cdot\boldsymbol{\mu}/\lambda}\hat{y}\,D\hat{y}\right\rangle_{\xi}+O(\epsilon^2)\\
&=-\frac{4}{\lambda\sqrt{2\pi}}\langle e^{-(1/2)(\boldsymbol{\xi}\cdot\boldsymbol{\mu}/\lambda)^2}\rangle_{\xi}+O(\epsilon^2)\ .\end{aligned} \tag{63}$$

These equations are to be used along with the relationship between λ and ρ, to define the $q=1$ line in the $(\mathbf{m},r)$ plane. However, for more than one condensed pattern the average over ξ makes the expression for $\mathbf{m}$ noninvertible, and the relation between λ and ρ is complicated for general $\mathbf{A}$.

Since $\tilde{\lambda}$ and ρ diverge faster than the term containing the integration variable y as $q\to 1$, the noise distribution becomes

$$\begin{aligned}\mathcal{D}_{\xi}[z^s,z^a]=\frac{\exp(-\frac{1}{2}z^a\mathcal{G}^{-1}z^a)}{\sqrt{8\pi^2\mathcal{G}\mathcal{E}}}&\left[\exp\left[-\frac{(z^s-\mathcal{B}+\Delta)^2}{4\mathcal{E}}\right]\Theta\left[-\frac{1}{\epsilon}\sum_{\nu\leq c}\xi^{\nu}\mu^{\nu}-(z^s+\Delta-\mathcal{B})\frac{\Gamma^2}{2\epsilon\mathcal{E}\rho}\right]\right.\\
&\left.+\exp\left[-\frac{(z^s-\mathcal{B}-\Delta)^2}{4\mathcal{E}}\right]\Theta\left[\frac{1}{\epsilon}\sum_{\nu\leq c}\xi^{\nu}\mu^{\nu}+(z^s-\Delta-\mathcal{B})\frac{\Gamma^2}{2\epsilon\mathcal{E}\rho}\right]\right],\end{aligned} \tag{64}$$

where $\epsilon=1-q$ and $\mathcal{G},\mathcal{E},\mathcal{B},\rho,\Delta$ are evaluated at $q=1$. Note, however, that the above analysis extends the noise distribution into the region where replica symmetry is unstable.

5. *de Almeida–Thouless (AT) surface*

The AT surface [17] signals the instability of the replica-symmetric solution via bifurcations of the form

$$q_{\alpha\beta}\to q+\delta q_{\alpha\beta},\quad \hat{q}_{ab}\to\frac{i}{2}\lambda^2+i\delta\hat{q}_{\alpha\beta}\ , \tag{65}$$

with $\sum_{\alpha\neq\beta}\delta q_{\alpha\beta}=0$, $\sum_{\alpha\neq\beta}\delta\hat{q}_{\alpha\beta}=0$ (the so-called replicon model). Using (27) we can expand about the replica symmetric solution in first nonvanishing orders giving

$$\Psi-\Psi_{\mathrm{RS}}\simeq\frac{1}{2}\sum_{\alpha\neq\beta}\sum_{\gamma\neq\delta}\delta q_{\alpha\beta}\delta q_{\gamma\delta}\frac{\partial^2\Omega}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}}+\frac{1}{2}\sum_{\alpha\neq\beta}\sum_{\gamma\neq\delta}\delta\hat{q}_{\alpha\beta}\delta\hat{q}_{\gamma\delta}\frac{\partial^2\Phi}{\partial\hat{q}_{\alpha\beta}\partial\hat{q}_{\gamma\delta}}-\sum_{\alpha\neq\beta}\delta q_{\alpha\beta}\delta\hat{q}_{\alpha\beta}\ , \tag{66}$$

where

$$\begin{aligned}\Omega&=\frac{1}{N}\ln\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})\exp(\tfrac{1}{2}\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\,\mathbf{m}\cdot\mathbf{A}^s\mathbf{z})\Phi\\
&=\ln\left\langle\left\langle\exp\left[\boldsymbol{\xi}\cdot\boldsymbol{\mu}\sum_{\alpha}s^{\alpha}+\frac{1}{2}\sum_{\alpha\neq\beta}[\lambda^2+2\delta\hat{q}_{\alpha\beta})s^{\alpha}s^{b}\right]\right\rangle_{\xi}\right\rangle_{\{s^{\alpha}\}}\ .\end{aligned} \tag{67}$$

The stability of the RS solution can be written as an eigenvalues problem,

$$\begin{bmatrix}\partial^2\Omega & -1\\ -1 & \partial^2\Phi\end{bmatrix}\begin{bmatrix}\delta q_{\alpha\beta}\\ \delta\hat{q}_{\alpha\beta}\end{bmatrix}=\Lambda\begin{bmatrix}\delta q_{\alpha\beta}\\ \delta\hat{q}_{\alpha\beta}\end{bmatrix}\ . \tag{68}$$

The critical stability will occur when the eigenvalue Λ is zero, so the AT surface defines a solution of $\partial^2\Omega\partial^2\Phi=1$. We leave the details of the calculation to Appendix B, where it is shown that the AT surface is given by

$$1=\frac{\rho^2}{N}\left[\frac{\rho^2}{p}(\mathrm{Tr}\tilde{\mathcal{R}})^2+2\rho\,\mathrm{Tr}(\mathbf{A}^s\tilde{\mathcal{R}})+\mathrm{Tr}(\mathbf{A}^s\mathbf{A}^s)\right]\left\langle\int D\hat{y}\cosh^{-4}(\xi\cdot\boldsymbol{\mu}+\lambda\hat{y})\right\rangle_{\xi} . \tag{69}$$

($\mathrm{Tr}\tilde{\mathcal{R}}=N\Delta$.) Replica symmetry is stable as long as the right hand side is greater than 1.

IV. REPLICA-SYMMETRIC ORDER PARAMETER FLOW

Having derived an expression for the noise distribution we now obtain differential equations for the flow of the order parameters $(\mathbf{m},r)$ by inserting (55) into (11). Upon making specific choices for the matrix $\mathbf{A}$ we can test the validity of our theory by comparing results from solving the macroscopic equations with results of performing microscopic spin simulations. As pointed out in [6,16] the present theory appears not to be exact due to the assumption of equipartitioning within the energy subshells. However, as with the previous study [3], we expect the equations to capture the essential features of the flow.

We can write the dynamic equations in a more compact form by changing variables to

$$\mathfrak{z}^a=\mathcal{G}^{-1/2}z^a\ ,\quad x'=\frac{z^s-\mathcal{B}\pm\Delta}{(2\mathcal{E})^{1/2}} . \tag{70}$$

The flow equations then become

$$\frac{d\mathbf{m}}{dt}=\left\langle\int D\mathfrak{z}^a\int Dx'\int Dy'\xi M_{\xi}(\mathfrak{z}^a,x',y')\right\rangle_{\xi}-\mathbf{m}\ , \tag{71}$$

$$\frac{1}{2}\frac{dr}{dt}=\left\langle\int D\mathfrak{z}^a\int Dx'\int Dy'R_{\xi}(\mathfrak{z}^a,x',y')\right\rangle_{\xi}-\left[r-\frac{1}{p}\mathrm{Tr}\,\mathbf{A}\right] , \tag{72}$$

where

$$\begin{aligned}M_{\xi}(\mathfrak{z}^a,x',y')=&\frac{1}{2}\left\{1-\tanh\left[\xi\cdot\boldsymbol{\mu}+y'\left[\frac{\Gamma^2\Delta}{2\mathcal{E}\rho}\right]^{1/2}+x'\frac{\Gamma^2}{(2\mathcal{E})^{1/2}\rho}\right]\right\}\tanh\beta(\xi\cdot\mathbf{Am}+\mathcal{G}^{1/2}\mathfrak{z}^a+U^-)\\&+\frac{1}{2}\left\{1+\tanh\left[\xi\cdot\boldsymbol{\mu}+y'\left[\frac{\Gamma^2\Delta}{2\mathcal{E}\rho}\right]^{1/2}+x'\frac{\Gamma^2}{(2\mathcal{E})^{1/2}\rho}\right]\right\}\tanh\beta(\xi\cdot\mathbf{Am}+\mathcal{G}^{1/2}\mathfrak{z}^a+U^+)\ ,\end{aligned} \tag{73}$$

$$\begin{aligned}R_{\xi}(\mathfrak{z}^a,x',y')=&\frac{1}{2}\left\{1-\tanh\left[\xi\cdot\boldsymbol{\mu}+y'\left[\frac{\Gamma^2\Delta}{2\mathcal{E}\rho}\right]^{1/2}+x'\frac{\Gamma^2}{(2\mathcal{E})^{1/2}\rho}\right]\right\}U^-\tanh\beta(\xi\cdot\mathbf{Am}+\mathcal{G}^{1/2}\mathfrak{z}^a+U^-)\\&+\frac{1}{2}\left\{1+\tanh\left[\xi\cdot\boldsymbol{\mu}+y'\left[\frac{\Gamma^2\Delta}{2\mathcal{E}\rho}\right]^{1/2}+x'\frac{\Gamma^2}{(2\mathcal{E})^{1/2}\rho}\right]\right\}U^+\tanh\beta(\xi\cdot\mathbf{Am}+\mathcal{G}^{1/2}\mathfrak{z}^a+U^+)\ ,\end{aligned} \tag{74}$$

$$U^{\pm}=(2\mathcal{E})^{1/2}x'+\mathcal{B}\mp\Delta\ . \tag{75}$$

A. Comparison with numerical simulations

We make the following specific choice for $\mathbf{A}$,

$$\mathbf{A}=\begin{bmatrix}\begin{smallmatrix}1&a\\-a&1\end{smallmatrix}&&&\mathbf{0}\\&\begin{smallmatrix}1&a\\-a&1\end{smallmatrix}&&\\&&\ddots&\\\mathbf{0}&&&\begin{smallmatrix}1&a\\-a&1\end{smallmatrix}\end{bmatrix}, \tag{76}$$

and assume that the first two patterns are condensed. The choice (76), without condensed-uncondensed coupling terms, is made purely for computational convenience. Without such terms the saddle points have unique solutions; if such terms were included they could become much more difficult to solve (increasing the already large amount of computer time necessary). By varying a we are still in a position to investigate the agreement between theory and simulations for both symmetric and asymmetric $\mathbf{A}$.

We notice that for $a=0$ we obtain the Hopfield model, and as $a\to\infty$ the interactions become predominantly antisymmetric. Because the saddle point equations contain only the symmetric part of the matrix, which is in this case the unit matrix, they simplify considerably,

$$\begin{bmatrix}m^1\\m^2\end{bmatrix}=\frac{1}{2}\int D\hat{y}\left[\begin{bmatrix}1\\1\end{bmatrix}\tanh(\mu^1+\mu^2+\lambda\hat{y})+\begin{bmatrix}1\\-1\end{bmatrix}\tanh(\mu^1-\mu^2+\lambda\hat{y})\right] , \tag{77}$$

$$q=\tfrac{1}{2}\int D\hat{y}\,[\tanh^2(\mu^1+\mu^2+\lambda\hat{y})+\tanh^2(\mu^1-\mu^2+\lambda\hat{y})]\ , \tag{78}$$

$$\frac{\lambda^2}{\alpha\rho^2}=\frac{q}{[1-\rho(1-q)]^2}\ ,\quad \tilde{r}=r=\frac{1-\rho(1-q)^2}{[1-\rho(1-q)]^2}\ ,\qquad \mathcal{G}=\alpha a^2 r\ . \tag{79}$$

The absence of coupling between condensed and uncondensed patterns leads to $\mathcal{B}=0$, $\Delta=[\alpha\rho(1-q)]/[1-\rho(1-q)]$, $2\mathcal{E}=\alpha r$, and $\mathbf{Am}=\binom{m^1+am^2}{-am^1+m^2}$. From these expressions the dynamics of the order parameters $(\mathbf{m},r)$ according to (72) can be calculated numerically.

In Figs. 3–5 we show the results of Monte Carlo spin simulations, along with the results of the theory in the m^1-m^2 and m^1-r planes, comprising flow lines of the simulations along with arrows indicating the theoretically predicted magnitude and direction of the derivatives dm^1/dm^2 and dm^1/dr. Each simulation is carried out for system of 40 000 spins, with each of the following sets of parameters: $a=0,1,2$; $T=0,1$; $p=1024,2048$.

As we can see the arrows give a good qualitative description of the direction of the actual flow. Closer analysis of the time derivatives in Figs. 6 and 7, however, shows deviations from the simulations that are noticeable for states giving rise to nonretrieval (i.e., flowing towards the fixed point $\mathbf{m}=0$, $r>1$), manifested in an overall slowing effect. The derivatives dm_2/dm_1 and dr/dm_1, however, show a better fit. It is not clear whether this is due to the effects for replica symmetry breaking or a manifestation of a more fundamental error within the formulation of the model. Such measurements were made for the Hopfield model [16], where small but significant discrepancies were found between the shape of the noise distribution found from the model and from simulations. Such effects were also observed in a toy model where an exact solution [4] was possible.

B. Fixed points of the flow

In the absence of asymmetric interactions we expect the fixed points of the flow (72) to correspond to the equilibrium saddle point equations derived in [13]. In order to show that this is indeed the case we carry out the following coordinate changes on the flow equations:

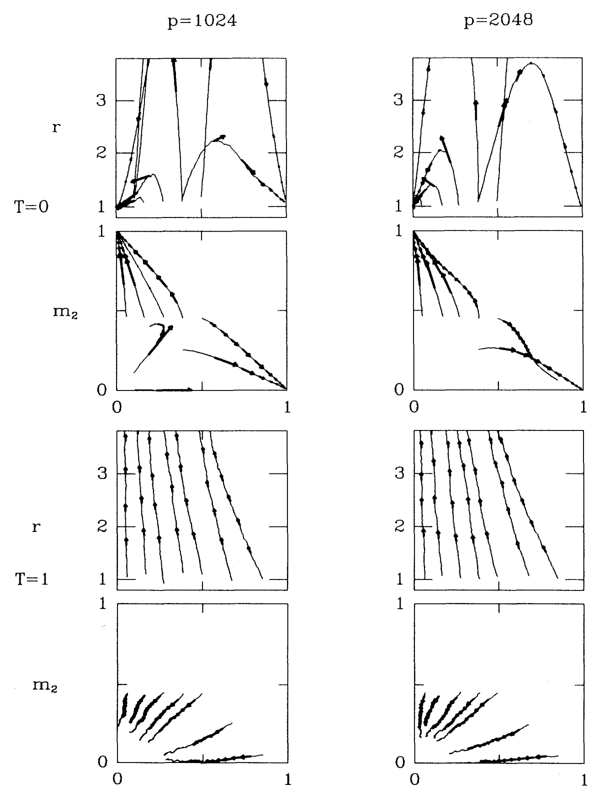

FIG. 3. Plots showing the evolution of m_1, m_2, and r for a system of 40 000 spins and embedding matrix (76) with $a=0$. The solid lines are the results of spin simulations, and the arrows show the instantaneous values of dm_2/dm_1 and dr/dm_1 calculated at intervals of $\frac{1}{2}$ iteration per spin. The lengths of the arrows are proportional to the magnitude of the derivatives.

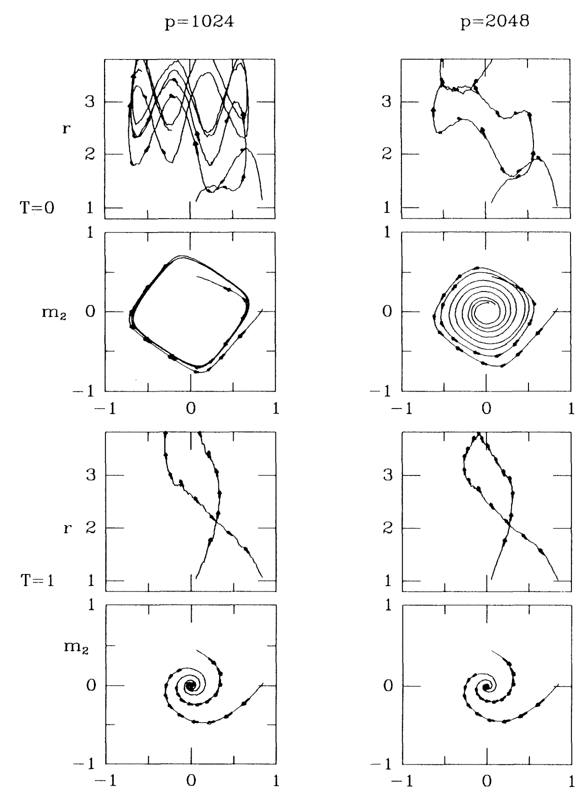

FIG. 4. Plots showing the evolution of m_1, m_2, and r for a system of 40 000 spins and embedding matrix (76) with $q=1$. The solid lines are the results of spin simulations, and the arrows show the instantaneous values of dm_2/dm_1 and dr/dm_1 calculated at intervals of $\frac{1}{2}$ iteration per spin. The lengths of the arrows are proportional to the magnitude of the derivatives.

$$x=x'\left[\frac{\Gamma^2}{2\mathcal{E}\rho^2}\right]^{1/2}+y'\left[\frac{\Delta}{2\mathcal{E}\rho}\right]^{1/2},$$
$$y=y'\left[\frac{\Gamma^2}{2\mathcal{E}\rho^2}\right]^{1/2}-x'\left[\frac{\Delta}{2\mathcal{E}\rho}\right]^{1/2}. \tag{80}$$

The flow equations then take the form

$$\frac{d\mathbf{m}}{dt}=\left\langle \int Dx\int Dy\,\xi M_\xi(\mathbf{m},r,x,y)\right\rangle_\xi-\mathbf{m},$$
$$\frac{1}{2}\frac{dr}{dt}=\left\langle \int Dx\int Dy\,R_\xi(\mathbf{m},r,xy)\right\rangle_\xi-\left[r-\frac{1}{p}\mathrm{Tr}\,\mathbf{A}\right], \tag{81}$$

where

$$M_\xi(\mathbf{m},r,x,y)=\tfrac{1}{2}[1-\tanh(\xi\cdot\boldsymbol{\mu}+\Gamma x)]\tanh(\beta\xi\cdot\mathbf{Am}+\beta U^-)+\tfrac{1}{2}[1+\tanh(\xi\cdot\boldsymbol{\mu}+\Gamma x)]\tanh(\beta\xi\cdot\mathbf{Am}+\beta U^+), \tag{82}$$

$$R_\xi(\mathbf{m},r,x,y)=\tfrac{1}{2}[1-\tanh(\xi\cdot\boldsymbol{\mu}+\Gamma x)]U^-\tanh(\beta\xi\cdot\mathbf{Am}+\beta U^-)+\tfrac{1}{2}[1+\tanh(\xi\cdot\boldsymbol{\mu}+\Gamma x)]U^+\tanh(\beta\xi\cdot\mathbf{Am}+\beta U^+), \tag{83}$$

$$U^\pm=\frac{\Gamma}{\rho}x-\left[\frac{\Delta}{\rho}\right]^{1/2}y-\mathcal{B}\mp\Delta. \tag{84}$$

We now use the two identities

$$\tanh u=\tfrac{1}{2}(1-\tanh u)\int Dy\,\tanh(u+yz-z^2)+\tfrac{1}{2}(1+\tanh u)\int Dy\,\tanh(u+yz+z^2), \tag{85}$$

$$u\tanh u+z^2=\tfrac{1}{2}(1-\tanh u)\int Dy(u+yz-z^2)\tanh(u+yz-z^2)+\tfrac{1}{2}(1+\tanh u)\int Dy(u+yz+z^2)\tanh(u+yz+z^2), \tag{86}$$

with $u=\xi\cdot\boldsymbol{\mu}+\Gamma x$, $z^2=\beta\Delta$, to show upon choosing $\rho=\beta$ and $\xi\cdot\boldsymbol{\mu}=\beta\xi\cdot\mathbf{Am}-\beta\mathcal{B}$ that the flow (81) is indeed at a fixed point

$$\frac{d\mathbf{m}}{dt}=\left\langle\int Dx\,\xi\tanh(\xi\cdot\boldsymbol{\mu}+\Gamma x)\right\rangle_\xi-\mathbf{m}=0, \tag{87}$$

$$\begin{aligned}\frac{1}{2}\frac{dr}{dt}&=\left\langle\int Dx\left[\frac{\xi\cdot\boldsymbol{\mu}}{\beta}+\frac{\Gamma x}{\beta}-\xi\cdot\mathbf{Am}\right]\tanh(\xi\cdot\boldsymbol{\mu}+\Gamma x)\right\rangle_\xi+\frac{\Delta}{\alpha}-\left[r-\frac{1}{p}\mathrm{Tr}\,\mathbf{A}\right]\\&=\frac{1}{\alpha\beta}\mathbf{m}\cdot(\boldsymbol{\mu}-\beta\,\mathbf{Am})+\frac{\Gamma^2}{\alpha\beta}\left[1-\left\langle\int Dx\,\tanh^2(\xi\boldsymbol{\mu}+\Gamma x)\right\rangle_\xi\right]+\frac{\Delta}{\alpha}-\left[r-\frac{1}{p}\mathrm{Tr}\,\mathbf{A}\right]\\&=\frac{1}{\alpha\beta}\mathbf{m}\cdot(\boldsymbol{\mu}-\beta\,\mathbf{Am})+\frac{\Gamma^2}{\alpha\beta}(1-q)+\frac{\Delta}{\alpha}-\left[r-\frac{1}{p}\mathrm{Tr}\,\mathbf{A}\right]=0,\end{aligned} \tag{88}$$

where the equalities are consequences of the saddle point equations (40). By comparing the dynamic saddle point equations with the equilibrium saddle point equations we can see that the equilibrium equations derived in [13] correspond to

$$\begin{aligned}\boldsymbol{\mu}&=\beta\{\mathbf{A}_{cc}^{-1}-\mathbf{A}_{cu}^{-1}[\mathbf{A}_{uu}^{-1}-\beta(1-q)\mathbb{1}]^{-1}\mathbf{A}_{uc}^{-1}\}^{-1}\mathbf{m}\\&=\beta\,\mathbf{A}_{uc}^{s}\mathbf{m}-\beta\rho(1-q)\mathbf{A}_{cu}^{s}[\mathbb{1}-\rho(1-q)\mathbf{A}_{uu}^{s}]^{-1}\mathbf{A}_{uc}^{s}\mathbf{m},\end{aligned} \tag{89}$$

$$\rho=\beta. \tag{90}$$

Careful analysis of the equilibrium saddle point equations reveal that these indeed imply $\xi\cdot\boldsymbol{\mu}=\beta\xi\cdot\mathbf{Am}-\beta\mathcal{B}$. Hence the present dynamic formalism correctly recovers the equilibrium phase diagram of [13] as fixed points of the flow.

V. CONCLUSIONS

In this paper we have generalized a recent theory [2,3] to describe the dynamics of the Hopfield model [1] near saturation to systems (i) with arbitrary separable interactions, (ii) which need not be symmetric, and (iii) with more than one condensed pattern. The theory, valid within the condensed ansatz, describes the evolution of macroscopic order parameters: the condensed overlaps m^μ $\mu=1,\ldots,c$ (where c is the number of condensed overlaps) and the disordered contribution to the energy r. The theory is based on the systematic removal of microscopic memory effects and requires the two assumptions of self-averaging with respect to the microscopic realizations of the stored patterns and equipartitioning of probability within the macroscopic subshells of the ensemble. While these assumptions are correct in detailed balance

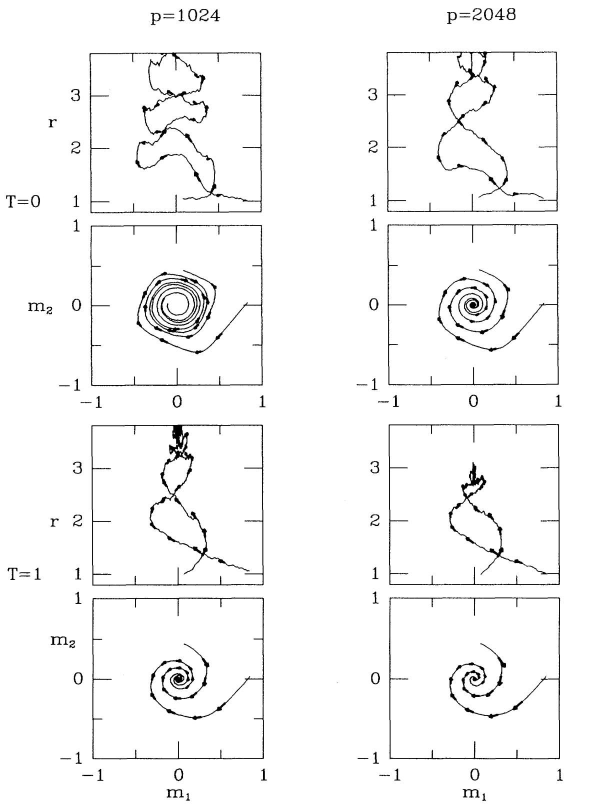

FIG. 5. Plots showing the evolution of m_1, m_2, and r for a system of 40 000 spins and embedding matrix (76) with $a=2$. The solid lines are the results of spin simulations, and the arrows show the instantaneous values of dm_2/dm_1 and dr/dm_1 calculated at intervals of $\frac{1}{2}$ iteration per spin. The lengths of the arrows are proportional to the magnitude of the derivatives.

equilibrium, there is no reason *a priori* to believe they will carry over to nonequilibrium, nonsymmetric cases. Indeed the results of previous studies [3,6,16] suggest that the assumption of equipartitioning within the $(\mathbf{m},r)$ subshells is not always valid, leading to an overall slowing down effect in the flows that leads to nonretrieval states ($\mathbf{m}=\mathbf{0}$). The closure of the dynamic laws requires the calculation of an intrinsic noise distribution, which can be calculated using the replica method. We have computed the region over which the replica-symmetric solution is stable. We have shown that the theory captures the essential characteristics of the flow for both symmetric and asymmetric interactions, while close observation of the flows and the noise distribution itself [16] show that the agreement between experiment and theory is not perfect.

Furthermore, in those regions where detailed balance holds, the correct equilibrium equations derived in [13] are recovered as stable fixed points of the dynamics. The theory has the additional advantage that the noise distribution can be calculated exactly within the limitations of the theory, without having to make *ad hoc* assumptions such as the Gaussian approximation, which is clearly not true in the regions where $q\to 1$. Our results imply that the present theory provides the first systematic method for analyzing nonsymmetric attractor neural networks near saturation, which cannot be analyzed within equilibrium statistical mechanics due to the absence of detailed balance.

ACKNOWLEDGMENT

We acknowledge financial support from the EPSRC under Grant Nos. 92310910 and GRH73028.

APPENDIX A: CALCULATION OF THE INTENSIVE CONTRIBUTION $\mathcal{R}$ TO THE INTRINSIC NOISE DISTRIBUTION

In this appendix we carry out some of the manipulations necessary to calculate the intensive contribution $\mathcal{R}$ (24) to the intrinsic noise distribution. We first define

$$g^{\mu}_{\alpha}=\frac{\int dz\, e^{-(1/2)\mathbf{z}\cdot\Lambda\mathbf{z}-i\Upsilon\cdot\mathbf{z}}z^{\mu}_{\alpha}}{\int d\mathbf{z}\, e^{-(1/2)\mathbf{z}\cdot\Lambda\mathbf{z}-i\Upsilon\cdot\mathbf{z}}}=i\frac{\partial}{\partial\Upsilon^{\mu}_{\alpha}}\ln\left[\int d\mathbf{z}\, e^{-(1/2)\mathbf{z}\cdot\Lambda\mathbf{z}-i\Upsilon\cdot\mathbf{z}}\right]=-i\sum_{\beta}\sum_{\nu}(\Lambda^{-1})^{\alpha\beta}_{\mu\nu}\Upsilon^{\nu}_{\beta}$$

$$=\rho\sqrt{N}\sum_{\beta}\sum_{\nu}(\Lambda^{-1})^{\alpha\beta}_{\mu\nu}A^{s}_{\nu\eta}m^{\eta}+O\left[\frac{1}{N}\right], \tag{A1}$$

$$g^{\mu\nu}_{\alpha\beta}=\frac{\int d\mathbf{z}\, e^{-(1/2)\mathbf{z}\cdot\Lambda\mathbf{z}-i\Upsilon\cdot\mathbf{z}}z^{\mu}_{\alpha}z^{\nu}_{\beta}}{\int d\mathbf{z}\, e^{-(1/2)\mathbf{z}\cdot\Lambda\mathbf{z}-i\Upsilon\cdot\mathbf{z}}}=-\frac{\partial^2}{\partial\Upsilon^{\mu}_{\alpha}\partial\Upsilon^{\nu}_{\beta}}\ln\left[\int d\mathbf{z}\, e^{-(1/2)\mathbf{z}\cdot\Lambda\mathbf{z}-i\Upsilon\cdot\mathbf{z}}\right]+g^{\mu}_{\alpha}g^{\nu}_{\beta}$$

$$=(\Lambda^{-1})^{\alpha\beta}_{\mu\nu}-\left[\sum_{\gamma\eta}(\Lambda^{-1})^{\alpha\gamma}_{\mu\eta}\Upsilon^{\eta}_{\gamma}\right]\left[\sum_{\delta\rho}(\Lambda^{-1})^{\beta\delta}_{\nu\rho}\Upsilon^{\rho}_{\delta}\right]. \tag{A2}$$

We now define

$$\Gamma_{\mu}=\sum_{\alpha}\left[\sum_{\nu>c}\Xi^{\alpha}_{\mu\nu}z^{\nu}_{\alpha}+\sum_{\nu\le c}\Pi^{\alpha}_{\mu\nu}m^{\nu}(\mathbf{s}^{\alpha})\right],$$

where

$$\Xi^{\alpha}_{\mu\nu}=\frac{1}{\sqrt{N}}[(xA^{s}_{\mu\nu}+yA^{a}_{\mu\nu})\delta_{\alpha 1}+i\rho A^{s}_{\mu\nu}s^{\alpha}_{i}]\ ,\quad \mu,\nu>c\ , \tag{A3}$$

$$\Pi^{\alpha}_{\mu\nu}=[(xA^{s}_{\mu\nu}+yA^{a}_{\mu\nu})\delta_{\alpha 1}+i\rho A^{s}_{\mu\nu}s^{\alpha}_{i}]\ ,\quad \mu>c,\nu\le c\ . \tag{A4}$$

Then $\mathcal{R}$ (24) becomes

$$\begin{aligned}\mathcal{R}=&\,ix\rho\sum_{\beta}\sum_{\mu,\rho\le c}\sum_{\nu,\eta>c}\zeta^{\mu}A^{s}_{\mu\nu}(\Lambda^{-1})^{1\beta}_{\nu\eta}A^{s}_{\eta\rho}m^{\rho}+\tfrac{1}{2}x^{2}\sum_{\mu,\rho\le x}\sum_{\nu,\eta>c}\zeta^{\mu}A^{s}_{\mu\nu}(\Lambda^{-1})^{11}_{\nu\eta}A^{s}_{\eta\rho}\zeta^{\eta}\\ &-\tfrac{1}{2}y^{2}\sum_{\mu,\rho\le x}\sum_{\nu,\eta>c}\zeta^{\mu}A^{a}_{\mu\nu}(\Lambda^{-1})^{11}_{\nu\eta}A^{a}_{\eta\rho}\zeta^{\eta}\\ &+\tfrac{1}{2}\sum_{\mu>c}\{\Xi^{\alpha}_{\mu\nu}(\Lambda^{-1})^{\alpha\beta}_{\nu\eta}\Xi^{\beta}_{\mu\eta}+[-i\Xi^{\alpha}_{\mu\nu}(\Lambda^{-1})^{\alpha\gamma}_{\nu\lambda}\Upsilon^{\lambda}_{\gamma}+\Pi^{\gamma}_{\mu\lambda}m^{\lambda}][-i\Xi^{\beta}_{\mu\eta}(\Lambda^{-1})^{\eta\rho}_{\beta\sigma}\Upsilon^{\rho}_{\delta}+\Pi^{\delta}_{\mu\rho}m^{\rho}]\}\ .\end{aligned} \tag{A5}$$

The symmetric and antisymmetric terms decouple. We neglect terms of order $1/N$ and define auxiliary variables $\tilde{\lambda}$,

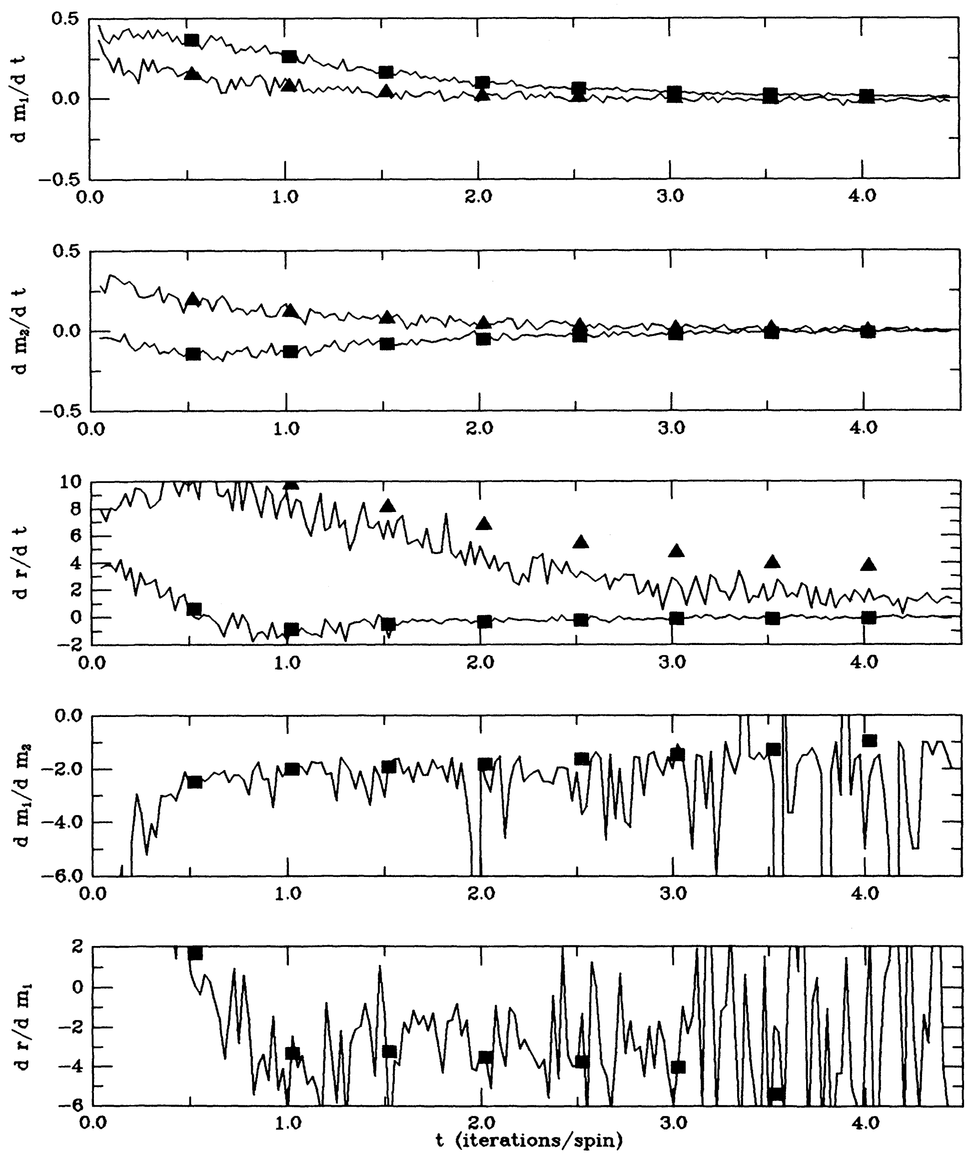

FIG. 6. Plots showing the derivatives dm_1/dt, dm_2/dt, dr/dt, dm_1/dm_2, and dr/dm_1 as a function of time measured in iterations per spin, for initial conditions leading to retrieval, i.e., $\mathbf{m}\neq\mathbf{0}$ (squares), and nonretrieval, i.e., $\mathbf{m}=\mathbf{0}$ (triangles). The lines show data taken from numerical simulations while the points are theoretical data. The system has 40 000 spins and embedding matrix (76) with $a=0$, $T=0$, and $p=1024$.

$\mathcal{A}_{\mu\nu}$, $\tilde{r}$, $\tilde{\mathcal{R}}_{\mu\nu}$, $\tilde{p}$, and $\tilde{\mathcal{P}}_{\mu\nu}$ such that

$$\frac{1}{N}\sum_{\mu\nu\lambda>c} A^{s}_{\lambda\mu}(\Lambda^{-1})^{\alpha\neq\beta}_{\mu\nu}A^{s}_{\nu\lambda}=\frac{\tilde{\lambda}^2}{\rho^2}=\frac{1}{N}\mathrm{Tr}\mathcal{A}\ ,\quad \frac{1}{N}\sum_{\mu\nu\lambda>c} A^{s}_{\lambda\mu}(\Lambda^{-1})^{\alpha=\beta}_{\mu\nu}A^{s}_{\nu\lambda}=\alpha\tilde{r}=\frac{1}{N}\mathrm{Tr}\tilde{\mathcal{R}}\ ,$$

$$\frac{1}{N}\sum_{\mu\nu\lambda>c} A^{a}_{\lambda\mu}(\Lambda^{-1})^{11}_{\mu\nu}A^{a}_{\nu\lambda}=\alpha\tilde{p}=\frac{1}{N}\mathrm{Tr}\tilde{\mathcal{P}}\ , \tag{A6}$$

where

$$\tilde{\lambda}^2=\frac{\alpha q\rho^2}{p}\mathrm{Tr}\{\,\mathbf{A}^{s}_{uu}\,\mathbf{A}^{s}_{uu}[\mathbb{1}-\rho(1-q+nq)\,\mathbf{A}^{s}_{uu}\,]^{-1}[\mathbb{1}-\rho(1-q)\,\mathbf{A}^{s}_{uu}\,]^{-1}\}\ , \tag{A7}$$

$$\tilde{r}_2=\frac{1}{p}\mathrm{Tr}\{\,\mathbf{A}^{s}_{uu}\,\mathbf{A}^{s}_{uu}[\mathbb{1}-\rho(1-q)(1-q+nq)\,\mathbf{A}^{s}_{uu}\,][\mathbb{1}-\rho(1-q+nq)\,\mathbf{A}^{s}_{uu}\,]^{-1}[\mathbb{1}-\rho(1-q)\,\mathbf{A}^{s}_{uu}\,]^{-1}\}\ . \tag{A8}$$

We can then simplify the expression for $\mathcal{R}$ to

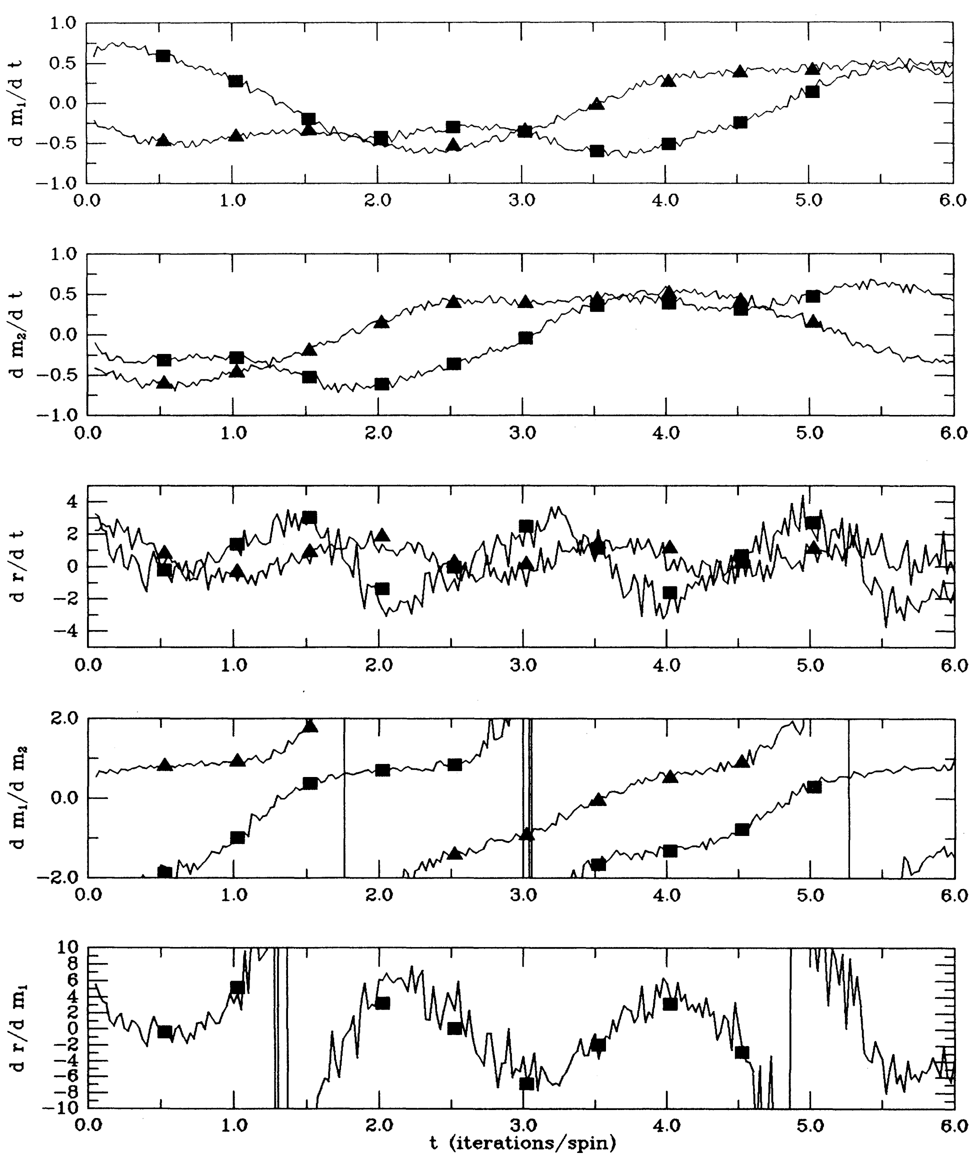

FIG. 7. Plots showing the derivatives dm_1/dt, dm_2/dt, dr/dt, dm_1/dm_2, and dr/dm_1 as a function of time measured in iterations per spin. The lines show data taken from numerical simulations while the points are theoretical data. The system has 40 000 spins and embedding matrix (76) with $a=1$, $T=0$, and $p=1024$ (squares) and $p=2048$ (triangles).

$$\mathcal{R}=ix\rho\boldsymbol{\zeta}\cdot[\tilde{\mathcal{R}}+(n-1)\lambda\mathcal{A}]\mathbf{m}+\frac{1}{2N}x^2\boldsymbol{\zeta}\cdot\tilde{\mathcal{R}}\boldsymbol{\zeta}-\frac{1}{2N}y^2\boldsymbol{\zeta}\cdot\tilde{\mathcal{P}}\boldsymbol{\zeta}+\frac{\alpha}{2}x^2\tilde{r}+ix\rho\left[\alpha\tilde{r}s_i^1+\frac{\tilde{\lambda}^2}{\rho^2}\sum_{\alpha>1}s_i^\alpha\right]$$
$$-\frac{\rho^2}{2}\left[\alpha\tilde{r}\sum_\alpha s_i^\alpha s_i^\alpha+\frac{\tilde{\lambda}^2}{\rho^2}\sum_{\alpha\neq\beta}s_i^\alpha s_i^\beta\right]-\frac{\alpha}{2}y^2\tilde{p}+\tfrac{1}{2}\rho^2x^2\mathbf{m}\cdot[(n-1)\lambda\mathcal{A}+\tilde{\mathcal{R}}][(n-1)\lambda\mathcal{A}+\tilde{\mathcal{R}}\mathbf{m}$$
$$+i\rho^3x\mathbf{m}\cdot[(n-1)\lambda\mathcal{A}+\tilde{\mathcal{R}}][(n-1)\lambda\mathcal{A}+\tilde{\mathcal{R}}]\mathbf{m}\sum_\alpha s_i^\alpha-\tfrac{1}{2}\rho^4\mathbf{m}\cdot[(n-1)\lambda\mathcal{A}+\tilde{\mathcal{R}}][(n-1)\lambda\mathcal{A}+\tilde{\mathcal{R}}]\mathbf{m}\left[\sum_\alpha s_i^\alpha\right]^2$$
$$-\tfrac{1}{2}y^2\rho^2\mathbf{m}\cdot\mathbf{A}^s_{cu}\tilde{\mathcal{P}}\mathbf{A}^s_{uc}\mathbf{m}-y^2\rho\mathbf{m}\cdot\sum_\beta\mathbf{A}^a_{cu}\mathbf{A}^a_{uu}(\Lambda^{-1})^{1\beta}\mathbf{A}^s_{uc}m+x^2\rho\mathbf{m}\cdot\mathbf{A}^s_{cu}[\tilde{\mathcal{R}}+(n-1)\lambda\mathcal{A}]\mathbf{m}$$
$$+i\rho^2x\mathbf{m}\cdot\mathbf{A}^s_{cu}\mathbf{A}^s_{uu}\left[\sum_{\alpha,\beta}s_i^\alpha(\Lambda^{-1})^{\alpha\beta}+\sum_{\gamma,\beta}s_i^c(\Lambda^{-1})^{1b}\right]\mathbf{A}^s_{uc}\mathbf{m}-\rho^3\mathbf{m}\cdot\mathbf{A}^s_{cu}\mathbf{A}^s_{uu}\sum_\gamma s_i^\gamma\sum_{\alpha,\beta}s_i^\alpha\Lambda^{-1\alpha\beta}\mathbf{A}^s_{uc}\mathbf{m}$$
$$+\tfrac{1}{2}x^2\mathbf{m}\cdot\mathbf{A}^s_{cu}\mathbf{A}^s_{uc}\mathbf{m}+i\rho x\mathbf{m}\cdot\mathbf{A}^s_{cu}\mathbf{A}^s_{uc}\mathbf{m}\sum_\alpha s_i^\alpha-\frac{\rho^2}{2}\mathbf{m}\cdot\mathbf{A}^s_{cu}\mathbf{A}^s_{uc}\mathbf{m}\sum_{\alpha,\beta}s_i^\alpha s_i^\beta-\tfrac{1}{2}y^2\mathbf{m}\cdot\mathbf{A}^a_{cu}\mathbf{A}^a_{uc}\mathbf{m}\ . \tag{A9}$$

We drop terms of $O(1/N)$ and hence the remaining part of the integral is of the form

$$\mathcal{D}_\zeta[z^s,z^a]\sim\int dx\,dy\,e^{ixz^s+iyz^a}e^{-(1/2)y^2\mathcal{G}}$$
$$\times\left\langle\exp\left[\sum_{\nu\le c}\xi_i^\nu\mu^\nu\sum_\alpha s_i^\alpha-ix\mathcal{B}-ix\rho\left[\alpha\tilde{r}s_i^1+\frac{\tilde{\lambda}^2}{\rho^2}\sum_{\alpha>1}s_i^\alpha\right]-x^2\mathcal{E}-i\rho x\sum_\alpha\mathcal{F}_\alpha s_i^\alpha\right.\right.$$
$$\left.\left.+\tfrac{1}{2}\Gamma^2\sum_{\alpha,\beta}s_i^\alpha s_i^\beta+\frac{n\rho^2\alpha\tilde{r}}{2}-\frac{n\tilde{\lambda}^2}{2}\right]\right\rangle_{\{s^\alpha\}}\ , \tag{A10}$$

where

$$\mathcal{G}=-\alpha\tilde{p}-2\rho\mathbf{m}\cdot\mathbf{A}^a_{cu}\mathbf{A}^a_{uu}\mathbf{A}^{s\ -1}_{uu}[\tilde{\mathcal{R}}+(n-1)\lambda\mathcal{A}]\mathbf{m}-\rho^2\mathbf{m}\cdot\mathbf{A}^s_{cu}\tilde{\mathcal{P}}\mathbf{A}^s_{uc}\mathbf{m}-\mathbf{m}\cdot\mathbf{A}^a_{cu}\mathbf{A}^a_{uc}\mathbf{m}\ , \tag{A11}$$

$$\mathcal{B}=\rho\boldsymbol{\zeta}\cdot[\tilde{\mathcal{R}}_{cc}+(n-1)\lambda\mathcal{A}_{cc}]\mathbf{m}\ , \tag{A12}$$

$$\mathcal{E}=\tfrac{1}{2}\mathbf{m}\cdot\{\mathbf{A}^s_{cu}+\rho[\tilde{\mathcal{R}}_{cu}+(n-1)\lambda\mathcal{A}_{cu}]\}\{\mathbf{A}^s_{uc}+\rho[\tilde{\mathcal{R}}_{uv}+(n-1)\lambda\mathcal{A}_{uc}]\}\mathbf{m}+\frac{\alpha\tilde{r}}{2}\ , \tag{A13}$$

$$\mathcal{F}_\alpha=\mathbf{m}\cdot\{\mathbf{A}^s_{cu}+\rho[\tilde{\mathcal{R}}_{cu}+(n-1)\lambda\mathcal{A}_{cu}]\}\{\mathbf{A}^s_{uc}+\rho[\tilde{\mathcal{R}}_{uc}+(n-1)\lambda\mathcal{A}_{uc}]\}\mathbf{m}\ , \tag{A14}$$

$$\Gamma^2=\tilde{\lambda}^2+\rho^2\mathbf{m}\cdot\{\mathbf{A}^s_{cu}+\rho[\tilde{\mathcal{R}}_{cu}+(n-1)\lambda\mathcal{A}_{cu}]\}\{\mathbf{A}^s_{uc}+\rho[\tilde{\mathcal{R}}_{uc}+(n-1)\lambda\mathcal{A}_{uc}]\}\mathbf{m}\ . \tag{A15}$$

Using our definition of Λ^{-1} we see that

$$\tilde{\mathcal{R}}+(n-1)\lambda\mathcal{A}=(1-q+nq)\mathbf{A}^s[\mathbb{1}-\rho(1-q)\mathbf{A}^s_{uu}]^{-1}\mathbf{A}^s\ , \tag{A16}$$

so that

$$\mathbf{m}\cdot\{\mathbf{A}^s_{cu}+\rho[\tilde{\mathcal{R}}_{cu}+(n-1)\lambda\mathcal{A}_{cu}]\}\{\mathbf{A}^s_{uc}+\rho[\tilde{\mathcal{R}}_{uc}+(n-1)\lambda\mathcal{A}_{uc}]\}\mathbf{m}=\mathbf{m}\cdot\mathbf{A}^s_{cu}[\mathbb{1}-\rho(1-q)\mathbf{A}^s_{uu}]^{-2}\mathbf{A}^s_{uc}\mathbf{m}\ , \tag{A17}$$

and therefore $\Gamma^2=\lambda^2$.

APPENDIX B: AT SURFACE

In this appendix we calculate the derivatives needed to determine the AT surface signaling the instability of the replica-symmetric solutions. To calculate the derivatives we note that

$$\int d\mathbf{z}\frac{\partial\mathcal{D}(\mathbf{z},\mathbf{q})}{\partial q_{\alpha\beta}}F(\mathbf{z})=\tfrac{1}{2}\sum_\mu\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})\frac{\partial^2F(\mathbf{z})}{\partial z^\mu_\alpha\partial z^\mu_\beta}\ . \tag{B1}$$

Therefore,

$$\frac{\partial^2\Omega}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}}$$

$$=\frac{\rho^2}{2N}\sum_{\mu}\frac{\partial}{\partial q_{\alpha\beta}}\frac{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}[(\mathbf{A}^s\mathbf{z}_\gamma)^\mu+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu][(\mathbf{A}^s\mathbf{z}_\delta)^\mu+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu]}{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}}$$

$$=\frac{\rho^2}{4N}\sum_{\mu\nu}\Bigg\{\frac{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})\frac{\partial^2}{\partial z_\alpha^\nu z_\beta^\nu}\{e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}[(\mathbf{A}^s\mathbf{z}_\gamma)^\mu+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu][(\mathbf{A}^s\mathbf{z}_\delta)^\mu+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu]\}}{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{a}\mathbf{A}^s\mathbf{z}}}$$

$$-\rho^2\left[\frac{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}[(\mathbf{A}^s\mathbf{z}_\alpha)^\nu+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\nu][(\mathbf{A}^s\mathbf{z}_\beta)^\nu+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\nu]}{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}}\right]$$

$$\times\left[\frac{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}[(\mathbf{A}^s\mathbf{z}_\gamma)^\mu+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu][(\mathbf{A}^s\mathbf{z}_\delta)^\mu+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu]}{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}}\right]\Bigg\}. \tag{B2}$$

Using previous notation we write

$$g_\alpha^\mu=\frac{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}z_\alpha^\mu}{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}}, \tag{B3}$$

$$g_{\alpha\beta}^{\mu\nu}=\frac{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}z_\alpha^\mu z_\beta^\nu}{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}}, \tag{B4}$$

$$g_{\alpha\beta\gamma}^{\mu\nu\eta}=\frac{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}z_\alpha^\mu z_\beta^\nu z_\gamma^\eta}{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}}, \tag{B5}$$

$$g_{\alpha\beta\gamma\delta}^{\mu\nu\eta\rho}=\frac{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}z_\alpha^\mu z_\beta^\nu z_\gamma^\eta z_\delta^\rho}{\int d\mathbf{z}\,\mathcal{D}(\mathbf{z},\mathbf{q})e^{(1/2)\rho\mathbf{z}\cdot\mathbf{A}^s\mathbf{z}+\rho\sqrt{N}\mathbf{m}\cdot\mathbf{A}^s\mathbf{z}}}. \tag{B6}$$

(B2) then becomes

$$\frac{\rho^2}{4N}\sum_{\mu\nu}\Bigg\{\rho^2\Bigg[\sum_{\alpha\beta\gamma\delta}A_{\mu\alpha}^sA_{\mu\beta}^sA_{\nu\gamma}^sA_{\nu\delta}^sg_{\alpha\beta\gamma\delta}^{\alpha\beta\gamma\delta}$$

$$+\sqrt{N}\sum_{\alpha\beta\gamma}[(\mathbf{m}\cdot\mathbf{A}^s)^\mu A_{\mu\alpha}^sA_{\nu\beta}^sA_{\nu\gamma}^s(g_{\alpha\beta\gamma}^{\alpha\beta\gamma}+g_{\alpha\beta\delta}^{\alpha\beta\gamma})+(\mathbf{m}\cdot\mathbf{A}^s)^\nu A_{\nu\alpha}^sA_{\mu\beta}^sA_{\mu\gamma}^s(g_{\alpha\gamma\delta}^{\alpha\beta\gamma}+g_{\beta\gamma\delta}^{\alpha\beta\gamma})]$$

$$+N\sum_{\alpha\beta}(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\mu A_{\nu\alpha}^sA_{\nu\beta}^sg_{\alpha\beta}^{\alpha\beta}$$

$$+\sum_{\alpha\beta}(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\nu(A_{\nu\alpha}^sA_{\mu\gamma}^sg_{\alpha\gamma}^{\alpha\gamma}+A_{\nu\alpha}^sA_{\mu\delta}^sg_{\alpha\delta}^{\alpha\delta}+A_{\nu\beta}^sA_{\mu\gamma}^sg_{\beta\gamma}^{\beta\gamma}+A_{\nu\beta}^sA_{\mu\delta}^sg_{\beta\delta}^{\beta\delta})$$

$$+(\mathbf{m}\cdot\mathbf{A}^s)^\nu(\mathbf{m}\cdot\mathbf{A}^s)^\nu A_{\mu\gamma}^sA_{\mu\delta}^sg_{\gamma\delta}^{\gamma\delta}$$

$$+N^{3/2}\sum_{\alpha}(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\nu[(\mathbf{m}\cdot\mathbf{A}^s)^\mu(A_{\nu\alpha}^sg_\alpha^\alpha+A_{\nu\beta}^sg_\beta^\beta)+(\mathbf{m}\cdot\mathbf{A}^s)^\nu(A_{\mu\gamma}^sg_\gamma^\gamma+A_{\mu\delta}^sg_\delta^\delta)]$$

$$+N^2(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\nu(\mathbf{m}\cdot\mathbf{A}^s)^\nu\Bigg]$$

$$+\rho A_{\mu\nu}^s\delta_{\beta\gamma}[A_{\nu\alpha}^sA_{\mu\delta}^sg_{\alpha\delta}^{\alpha\delta}+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\nu A_{\mu\delta}^sg_\delta^\delta+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu A_{\nu\alpha}^sg_\alpha^\alpha+N(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\nu]$$

$$+\rho A_{\mu\nu}^s\delta_{\beta\delta}[A_{\nu\alpha}^sA_{\mu\gamma}^sg_{\alpha\gamma}^{\alpha\gamma}+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\nu A_{\mu\gamma}^sg_\gamma^\gamma+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu A_{\nu\alpha}^sg_\alpha^\alpha+N(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\nu]$$

$$+\rho A_{\mu\nu}^s\delta_{\alpha\gamma}[A_{\nu\beta}^sA_{\mu\delta}^sg_{\beta\delta}^{\beta\delta}+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\nu A_{\mu\delta}^sg_\delta^\delta+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu A_{\nu\beta}^sg_\beta^\beta+N(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^2)^\nu]$$

$$+\rho A^s_{\mu\nu}\delta_{\alpha\delta}[A^s_{\nu\beta}A^s_{\mu\gamma}g^{\beta\gamma}_{\beta\gamma}+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\nu A^s_{\mu\gamma}g^\gamma_\gamma+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu A^s_{\nu\beta}g^\beta_\beta+N(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\nu]$$
$$+A^s_{\mu\nu}A^s_{\mu\nu}(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma})$$
$$-\rho^2[A^s_{\nu\alpha}A^s_{\nu\beta}g^{\alpha\beta}_{\alpha\beta}+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\nu A^s_{\nu\alpha}g^\alpha_\alpha+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\nu A^s_{\nu\alpha}g^\alpha_\beta+N(\mathbf{m}\cdot\mathbf{A}^s)^\nu(\mathbf{m}\cdot\mathbf{A}^s)^\nu]$$
$$\left.\times[A^s_{\mu\gamma}A^s_{\mu\delta}g^{\gamma\delta}_{\gamma\delta}+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu A^s_{\mu\alpha}g^\gamma_\gamma+\sqrt{N}(\mathbf{m}\cdot\mathbf{A}^s)^\mu A^s_{\mu\gamma}g^\delta_\delta+N(\mathbf{m}\cdot\mathbf{A}^s)^\mu(\mathbf{m}\cdot\mathbf{A}^s)^\mu]\right\} . \tag{B7}$$

We can calculate (114) using

$$-i\frac{\partial\Omega}{\partial\Upsilon^\mu_\alpha}=g^\mu_\alpha=\rho\sqrt{N}\sum_\beta\sum_\nu(\Lambda^{-1})^{\alpha\beta}_{\mu\nu}A^s_{\nu\eta}m^\eta ,$$
$$-\frac{\partial^2\Omega}{\partial\Upsilon^\mu_\alpha\partial\Upsilon^\nu_\beta}=g^{\mu\nu}_{\alpha\beta}-g^\mu_\alpha g^\nu_\beta=(\Lambda^{-1})^{\alpha\beta}_{\mu\nu} ,$$
$$i\frac{\partial^3\Omega}{\partial\Upsilon^\mu_\alpha\partial\Upsilon^\nu_\beta\partial\Upsilon^\eta_\gamma}=g^{\mu\nu\eta}_{\alpha\beta\gamma}-g^{\mu\nu}_{\alpha\beta}g^\eta_\gamma-g^{\mu\eta}_{\alpha\gamma}g^\nu_\beta-g^{\nu\eta}_{\beta\gamma}g^\mu_\alpha+2g^\mu_\alpha g^\nu_\beta g^\eta_\gamma=0 ,$$
$$\frac{\partial^4\Omega}{\partial\Upsilon^\mu_\alpha\partial\Upsilon^\nu_\beta\partial\Upsilon^\eta_\gamma\partial\Upsilon^\delta_d}=g^{\mu\nu\eta\delta}_{\alpha\beta\gamma\delta}-g^{\mu\nu\eta}_{\alpha\beta\gamma}g^\delta_d-g^{\mu\nu\delta}_{\alpha\beta\delta}g^\eta_\gamma-g^{\mu\nu}_{\alpha\beta}g^{\eta\delta}_{\gamma\delta}+2g^{\mu\nu}_{\alpha\beta}g^\eta_\gamma g^\delta_\delta-g^{\mu\eta\delta}_{\alpha\gamma\delta}g^\nu_\beta-g^{\mu\eta}_{\alpha\gamma}g^{\nu\delta}_{\beta\delta}+2g^{\mu\eta}_{\alpha\gamma}g^\nu_\beta g^\delta_\delta$$
$$-g^{\nu\eta\delta}_{\beta\gamma\delta}g^\mu_\alpha-g^{\mu\delta}_{\alpha\delta}g^{\nu\eta}_{\beta\gamma}+2g^{\nu\eta}_{\beta\delta}g^\mu_\alpha g^\delta_\delta+2g^{\mu\delta}_{\alpha\delta}g^\nu_\beta g^\eta_\gamma+2g^\mu_\alpha g^{\nu\delta}_{\beta\delta}g^\eta_\gamma+2g^\mu_\alpha g^\nu_\beta g^{\eta\delta}_{\gamma\delta}-6g^\mu_\alpha g^\nu_\beta g^\eta_\gamma g^\delta_\delta=0 ,$$
$$g^{\mu\nu\eta\delta}_{\alpha\beta\gamma\delta}=g^{\mu\nu}_{\alpha\beta}g^{\eta\delta}_{\gamma\delta}+g^{\mu\eta}_{\alpha\gamma}g^{\nu\delta}_{\beta\delta}+g^{\mu\delta}_{\alpha\delta}g^{\nu\eta}_{\beta\gamma} . \tag{B8}$$

Since the replicon mode has the property $\sum_\alpha\delta q_{\alpha\beta}=0$ [17], we need only consider terms containing exactly two δ functions (in replica space). Therefore the only important terms in the above expression are

$$\frac{\rho^2}{4N}(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma})\sum_{\mu\nu}\left\{\frac{\rho^2}{p}\left[\sum_{\alpha\beta\gamma\delta}A^s_{\mu\alpha}A^s_{\mu\beta}A^s_{\nu\gamma}A^s_{\nu\delta}\hat{g}^{\alpha\beta}\hat{g}^{\gamma\delta}\right]+2\rho\sum_{\alpha\beta}A^s_{\mu\nu}A^s_{\nu\alpha}A^s_{\mu\beta}\hat{g}^{\alpha\beta}+A^s_{\mu\nu}A^s_{\mu\nu}\right\} ,$$

where $\hat{g}^{\alpha\beta}=g^{\alpha\beta}_{\alpha\alpha}$.

Similarly we can show that the important terms in

$$\frac{\partial^2\Phi}{\partial\delta\hat{q}_{\alpha\beta}\partial\delta\hat{q}_{\gamma\delta}}=\frac{\partial}{\partial\hat{q}_{\alpha\beta}}\frac{\left\langle\left\langle s^\gamma s^\delta\exp\left[\xi\cdot\mu\sum_\alpha s^\alpha+\frac{1}{2}\sum_{\alpha\neq\beta}(\lambda^2+2\delta\hat{q}_{\alpha\beta})s^\alpha s^\beta\right]\right\rangle_\xi\right\rangle_{\{s^\alpha\}}}{\left\langle\left\langle\exp\left[\xi\cdot\mu\sum_\alpha s^\alpha+\frac{1}{2}\sum_{\alpha\neq\beta}(\lambda^2+2\delta\hat{q}_{\alpha\beta})s^\alpha s^\alpha\right]\right\rangle_\xi\right\rangle_{\{s^\alpha\}}} \tag{B9}$$

are

$$(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma})\left\langle\int D\hat{y}\cosh^{-4}(\xi\cdot\mu+\lambda\hat{y})\right\rangle_\xi . \tag{B10}$$

[1] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[2] A. C. C. Coolen and D. Sherrington, Phys. Rev. Lett. **71**, 3886 (1993).
[3] A. C. C. Coolen and D. Sherrington, Phys. Rev. E **36**, 1921 (1994).
[4] A. C. C. Coolen and S. Franz, J. Phys. A **27**, 6947 (1994).
[5] A. C. C. Coolen and D. Sherrington, J. Phys. A **27**, 7687 (1994).
[6] D. Sherington and A. C. C. Coolen, in *Proceedings of the Sites Conference on 25 Years of Nonequilibrium Statistical Mechanics,* edited by L. Garrido (Springer-Verlag, Berlin, to be published).
[7] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).
[8] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).
[9] J. L. van Hemmen and R. Kühn, in *Models of Neural Networks,* edited by E. Domany, J. L. van Hemmen, and K. Schulten (Springer-Verlag, Berlin, 1991), Chap. 1.
[10] H. Nishimori and T. Ozeki, J. Phys. A **27**, 7061 (1994).
[11] S. Amari and K. Maginu, Neural Networks **1**, 63 (1988).
[12] G. A. Kohring, J. Phys. A **23**, 2237 (1990).
[13] L. F. Cugliandolo and M. V. Tsodyks, J. Phys. A **27**, 741 (1994).
[14] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. **173**, 30 (1987).
[15] A. C. C. Coolen and T. W. Ruijgrok, Phys. Rev. A **38**, 4253 (1988).
[16] T. Ozeki and H. Nishimori, J. Phys. A **27**, 7061 (1994).
[17] J. R. L. de Almeida and D. J. Thouless, J. Phys. A **11**, 983 (1978).