Noise distributions in retrieval dynamics of the Hopfield model

1994 J. Phys. A: Math. Gen. 27 7061

(http://iopscience.iop.org/0305-4470/27/21/020)

J. Phys. A: Math. Gen. 27 (1994) 7061–7068. Printed in the UK

Noise distributions in retrieval dynamics of the Hopfield model

Tomoko Ozeki and Hidetoshi Nishimori
Department of Physics, Tokyo Institute of Technology, Oh-okayama, Meguro, Tokyo 152, Japan

Received 14 June 1994

Abstract. We examine the dynamical theory, recently proposed by Coolen and Sherrington, which describes the deterministic flow of order parameters in the memory retrieval processes of the Hopfield model. We have performed Monte Carlo simulations to investigate the limit of applicability of their theory. We have found that the qualitative behaviour of retrieval processes follows their prediction fairly well. However, the theory fails to describe the noise distribution quantitatively when the effects of non-retrieved patterns are not negligible.

Many attempts [1–9] have been made to describe the dynamics of retrieval processes in the Hopfield model. Great interest lies in such non-trivial behaviour as the sensitivity to initial conditions and the sizes and shapes of the basins of attraction. These properties cannot be determined without a direct treatment of the dynamical process. Despite its necessity, there is, as yet, no such exact dynamical theory available. As Gardner *et al* [4] suggested, the number of order parameters representing the network state increases very quickly with the time-steps and this makes it very difficult to handle the dynamical processes rigorously. Amari and Maginu [5] employed a signal-to-noise ratio analysis and tried to describe the dynamical processes approximately in terms of a few macrovariables. Although their theory is simple and a good approximation for successful retrieval processes, analysis in our previous paper [3] clearly showed that their assumption that the distribution of the noise term in the local field is Gaussian is not satisfied when retrieval fails. The noise distribution is non-Gaussian from an early stage of the memory retrieval process which leads the networks to a spin–glass (non-retrieval) state. Therefore an exact treatment of the noise distribution was required to describe the dynamical processes correctly when the retrieval process is not successful. Coolen and Sherrington [1, 2] (CS hereafter) recently proposed a dynamical theory to describe the retrieval processes analytically. They employed a replica calculation of the distribution of the noise term under certain assumptions on the microscopic behaviour of the system. The purpose of the present paper is to show the results of extensive numerical simulations which examine the limit of applicability of their theory. We have found that the predicted noise distribution agrees fairly well qualitatively with simulation results. Thus this theory is a step forward from the Amari–Maginu theory [5] which employs a simple Gaussian form of the noise distribution. However, in quantitative terms, deviations have been observed between the theory and simulations even in the region where CS conjectured the theory to be exact.

We consider the memory retrieval process of the Hopfield model. The Hopfield model consists of N neurons, each of which has a two-state variable $s_i = \pm 1$. Dynamical flow of the network state $\boldsymbol{s}$ is governed by an asynchronous stochastic updating rule

$$w_k(\boldsymbol{s}) \equiv \tfrac{1}{2}\{1 - s_k \tanh[\beta h_k(\boldsymbol{s})]\} \qquad (1)$$

0305-4470/94/217061+08$19.50

where $w_k(s)$ is the transition rate of the kth spin and $h_k(s)$ denotes the local field to the kth neuron in the state s:

$$h_k(s) = \sum_{j \neq k} J_{kj} s_j \,.$$

When we store the random patterns ξ_i^μ ($\mu = 1, \ldots, \alpha N$) according to the Hebb rule

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{\alpha N} \xi_i^\mu \xi_j^\mu \tag{2}$$

the embedded patterns become attractors of the dynamical flow of the network state under appropriate conditions. CS started from the microscopic equations of the Markov process, described by the master equation

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t(s) = \sum_{k=1}^{N} [P_t(F_k s) w_k(F_k s) - P_t(s) w_k(s)]$$

in which F_k is a spin–flip operator,

$$F_k \Phi(s) \equiv \Phi(s_1, \ldots, -s_k, \ldots, s_N)$$

and obtained the deterministic dynamical equations for order parameters

$$m(s) \equiv \frac{1}{N} \sum_{k=1}^{N} \xi_k^1 s_k$$

$$r(s) \equiv \frac{1}{\alpha} \sum_{\mu>1}^{\alpha N} \left[\frac{1}{N} \sum_{k=1}^{N} \xi_k^\mu s_k \right]^2 .$$

Here $m(s)$ is the overlap of the network state with the first pattern (supposing that the network is in the process where the first pattern ξ^1 is retrieved) and $r(s)$ represents the overlap of the network state with non-retrieved patterns.

The local field can be written as

$$h_i(s) = \xi_i^1 [m(s) + z_i(s)] - \frac{1}{N} s_i \qquad z_i(s) \equiv \xi_i^1 \sum_{\mu>1}^{\alpha N} \xi_i^\mu \frac{1}{N} \sum_{k \neq i} \xi_k^\mu s_k \tag{3}$$

where $z_i(s)$ is the so-called noise term in the local field. They introduced a distribution which measures the probability density in terms of the macroscopic order parameters (m, r)

$$P_t(m, r) = \sum_{s} P_t(s) \delta[m - m(s)] \delta[r - r(s)] \,.$$

By inserting the microscopic equation, they wrote the time derivative of this macroscopic distribution for large N satisfying the Liouville form. Then they obtained the deterministic evolution of the order parameters (m, r)

$$P_t(m, r) = \delta(m - m^*(t)) \delta(r - r^*(t))$$

in which the deterministic trajectory $(m^*(t), r^*(t))$ is given by the solution of the set of flow equations

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \int \mathrm{d}z \, D_{m,r;t}[z] \tanh[\beta m + \beta z] - m \tag{4}$$

$$\frac{1}{2} \frac{\mathrm{d}r}{\mathrm{d}t} = \frac{1}{\alpha} \int \mathrm{d}z \, D_{m,r;t}[z] z \tanh[\beta m + \beta z] + 1 - r \tag{5}$$

where

$$D_{m,r;t}[z] \equiv \frac{\sum_s P_t(s)\delta[m-m(s)]\delta[r-r(s)]\frac{1}{N}\sum_i \delta[z-z_i(s)]}{\sum_s P_t(s)\delta[m-m(s)]\delta[r-r(s)]} \tag{6}$$

is the noise distribution in the (m, r) subshells.

For simplicity of calculation of the noise distribution, CS assumed the following.

(i) The flow equations of the order parameters (4) and (5), and therefore the noise distribution $D_{m,r;t}[z]$ (6), are self-averaging in the thermodynamic limit ($N \to \infty$) with respect to pattern realizations $\{\xi^\mu\}$.

(ii) As far as the calculation of the noise distribution is concerned, CS assumed equipartitioning of probability in the macroscopic (m, r) subshells of the ensemble.

The former allowed them to simplify the calculation of the noise distribution by averaging over the random patterns $\{\xi^\mu\}$. The latter removed the explicit time dependence coming from $P_t(s)$ in the noise distribution.

$$D_{m,r;t}[z] \to D_{m,r}[z] \equiv \left\langle \frac{\sum_s \delta[m-m(s)]\delta[r-r(s)]\frac{1}{N}\sum_i \delta[z-z_i(s)]}{\sum_s \delta[m-m(s)]\delta[r-r(s)]} \right\rangle_{\{\xi\}} . \tag{7}$$

These two assumptions enabled them to perform the replica calculation of the noise distribution (details are found in [1]). The result under a replica-symmetric (RS) ansatz is

$$D^{\rm RS}_{m,r}[z] = \frac{e^{-(\Delta+z)^2/2\alpha r}}{2\sqrt{2\pi\alpha r}}\left\{1 - \int Dy \tanh\left[\lambda y\left[\frac{\Delta}{\alpha\rho r}\right]^{1/2} + (\Delta+z)\frac{\lambda^2}{\alpha\rho r} + \mu\right]\right\}$$
$$+ \frac{e^{-(\Delta-z)^2/2\alpha r}}{2\sqrt{2\pi\alpha r}}\left\{1 - \int Dy \tanh\left[\lambda y\left[\frac{\Delta}{\alpha\rho r}\right]^{1/2} + (\Delta-z)\frac{\lambda^2}{\alpha\rho r} - \mu\right]\right\}$$

with a set of saddle-point equations

$$q = \int Dy \tanh^2[\lambda(q)y + \mu] \qquad \lambda(q) = \frac{\sqrt{\alpha q}}{1-q}\,\frac{2r-1+q-\sqrt{(1-q)^2+4rq}}{1-q+\sqrt{(1-q)^2+4rq}}$$
$$m = \int Dy \tanh[\lambda y + \mu] \qquad \rho(q) = \frac{1}{2r(1-q)}[2r-1+q-\sqrt{(1-q)^2+4rq}]$$

where $\Delta \equiv \alpha\rho r - \lambda^2/\rho$. The Gaussian measure is denoted by $Dy = (2\pi)^{-1/2}e^{-y^2/2}\,dy$.

We have performed Monte Carlo simulations to check the limit of applicability of CS' assumptions on the noise distribution mentioned above. We have simulated the retrieval processes in the Hopfield network of asynchronous dynamics with $N = 5000, 10\,000$ and $30\,000$ neurons at zero temperature. We embedded αN random patterns ξ^μ in the network according to the Hebb rule (2). The storage capacity α was chosen to be 0.1 and 0.2 for each N. We generated initial states of the network by flipping each spin of the first pattern ξ_i^1 independently with probability $p = \frac{1}{2}(1 - m_0)$. Therefore, on average, our initial states have a finite overlap m_0 with the first pattern. The initial overlap m_0 was set to $0.1, 0.4$ and 0.8. At zero temperature (in the limit of $\beta \to \infty$), the transition rate (1) becomes

$$w_k(s) = \tfrac{1}{2}\{1 - s_k \mathrm{sgn}[h_k(s)]\} . \tag{8}$$

Starting from the initial state, we have examined the dynamical behaviour by selecting a single neuron randomly and flipping the neuron according to the transition rate (8). In the retrieval process, we have numerically calculated order parameters (m, r) and the distribution of the noise term $z_i(s)$ (3). We have also solved the set of saddle-point equations

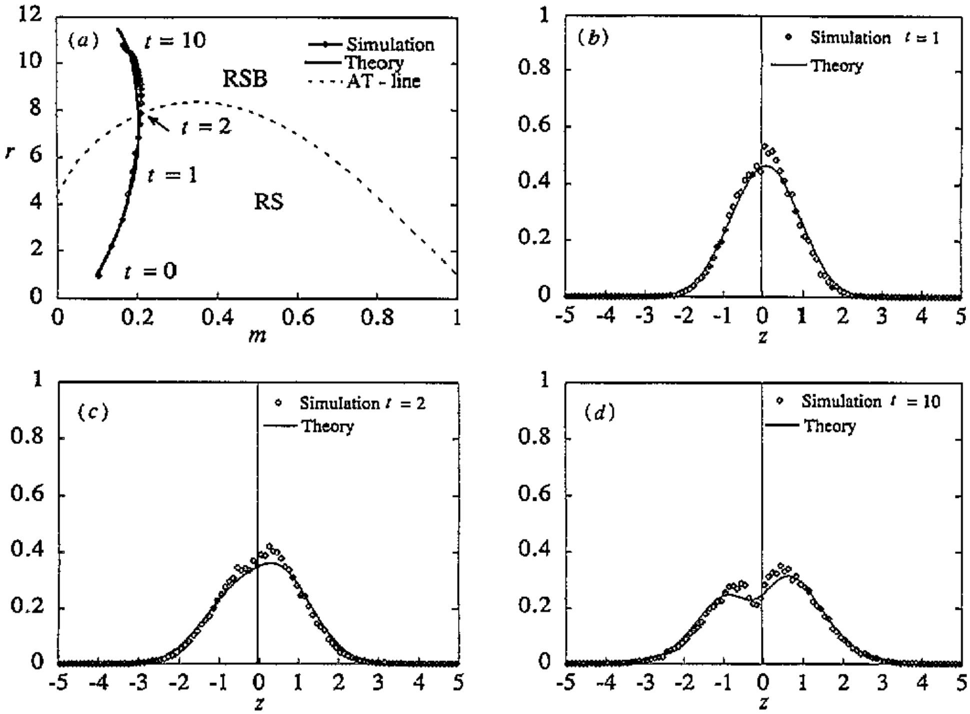

Figure 1. (a) Time evolution of the network in the m versus r diagram (non-successful retrieval); $T = 0, N = 30\,000, \alpha = 0.1$ and $m_0 = 0.1$ (likewise for (b)–(d)). Simulation results are plotted as $\Diamond$ at every $\frac{1}{4}$ iterations per neuron. The broken curve is the AT-line, above this the replica-symmetric solution is unstable. The full curve represents the CS prediction for the time evolution of the order parameters. (b) Noise distribution at $t = 1$ (iteration per neuron). The full curve was obtained for the RS distribution by CS. The network is in the RS region. (c) As for (b), with noise distribution at $t = 2$. (d) As for (c), with noise distribution at $t = 10$.

using (m, r) obtained by simulation and calculated the replica-symmetric noise distribution $D^{\rm RS}_{m,r}[z]$. We have compared the distribution of the noise term of simulations with that obtained by the CS theory $D^{\rm RS}_{m,r}[z]$ for the same (m, r).

One of the results is shown in figure 1, which represents the case $N = 30\,000, \alpha = 0.1$ and $m_0 = 0.1$. Figure 1(a) shows the development of the order parameters (m, r). The overlap m increases initially, but decreases after a while and finally retrieval fails. The initial overlap $m_0 = 0.1$ is too small for successful retrieval. We have also solved differential equations (4) and (5) numerically to compare the CS prediction for the time evolution of the order parameters (the full curve) with simulation results. The broken curve represents the replica-symmetry breaking (AT) curve

$$0 = \alpha - \rho^2(\alpha + \Delta)^2 \int \frac{{\rm D}y}{\cosh^4[\lambda y + \mu]} \,.$$

The RS calculation of noise distribution does not hold beyond this line in the m–r diagram. The noise distributions are depicted in figures 1(b)–(d). We have taken statistics of z_i by varying i from 1 to N at each time-step (iteration per neuron) for a fixed set of random patterns ξ^μ. The full curve is obtained by the RS calculation of the noise distribution $D^{\rm RS}_{m,r}[z]$. In figures 1(b) and (c), the network state is in the RS region. The CS theory describes the noise distribution qualitatively well. Compared to the Amari–Maginu theory

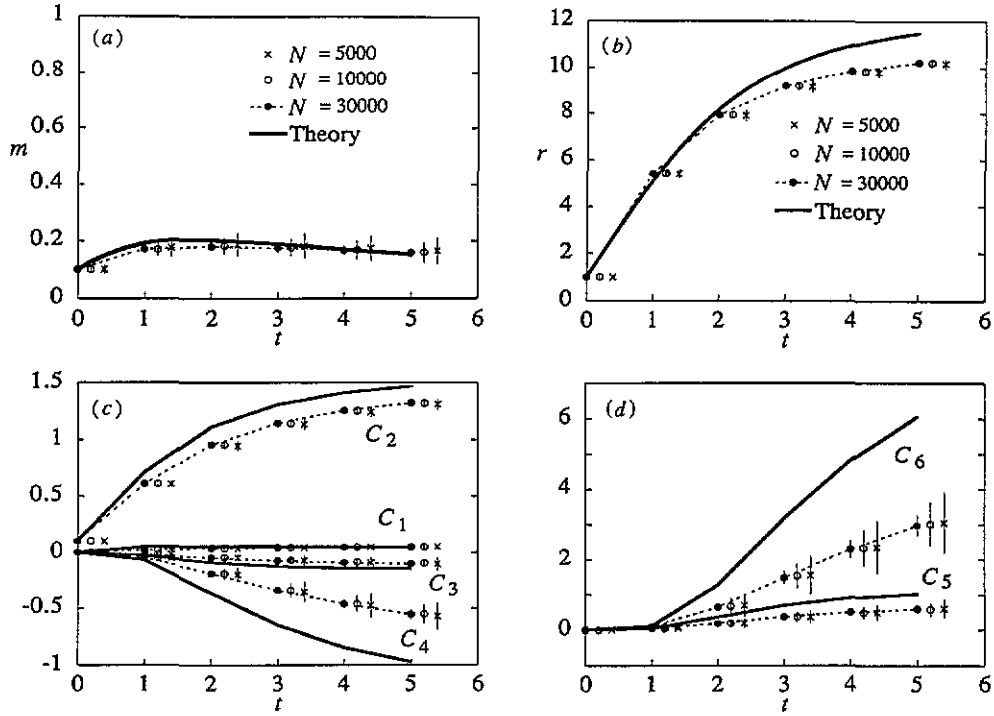

Figure 2. (*a*) Time-dependent behaviour of m (non-successful retrieval); $T = 0, \alpha = 0.1$ and $m_0 = 0.1$. ×, ○ and • denote simulation results for $N = 5000, 10\,000$ and $30\,000$, respectively. Plots were obtained by averaging over 100 samples, error bars represent standard deviations, and the broken curve is a guide for the eye. For $N = 5000$ and $10\,000$, each data point is shown shifted slightly in the horizontal direction to avoid overlaps. Self-averaging with respect to pattern realization is likely to hold. The full curve represents the CS prediction for the time evolution of the order parameter. (*b*) Time-dependent behaviour of r, obtained under the same condition as (*a*). Self-averaging with respect to pattern realization is likely to hold. (*c*) Time evolution of the cumulants of noise distribution to fourth order; conditions as in (*a*). C_1 to C_4 represent cumulants up to fourth order. Full curves denote CS' RS calculation. There exist clear differences between the theory and simulations even when the network is in the RS region ($t \leqslant 2$). (*d*) Time evolution of the fifth and sixth cumulants of noise distribution; conditions as in (*c*). (Full curves denote CS' RS calculation.) There exist differences between the theory and simulations even when the network is in the RS region ($t \leqslant 2$).

[5] which assumed a Gaussian distribution, the CS theory succeeds in depicting the noise distribution in more detail. However, deviations of simulation results from the theory are *apparently non-negligible quantitatively*.

As shown in our previous paper [3], cumulants of the noise distribution allow us to discuss the theory in a quantitative fashion. We have calculated cumulants of the distribution to sixth order. Moreover, to check the finite-size effect and CS' assumption of self-averaging property, we have investigated 100 different sets of random patterns for each $N = 5000, 10\,000$ and $30\,000$. The results are shown in figure 2. Simulation conditions are the same as in figure 1 and retrieval fails. Figures 2(*a*) and (*b*) show the time evolution of m and r. The cross, open and full circles show the simulation results for $N = 5000, 10\,000$ and $30\,000$, respectively. As CS suggested, the simulation results back up their assumption of the self-averaging property. As the network size increases, fluctuations decrease and for large

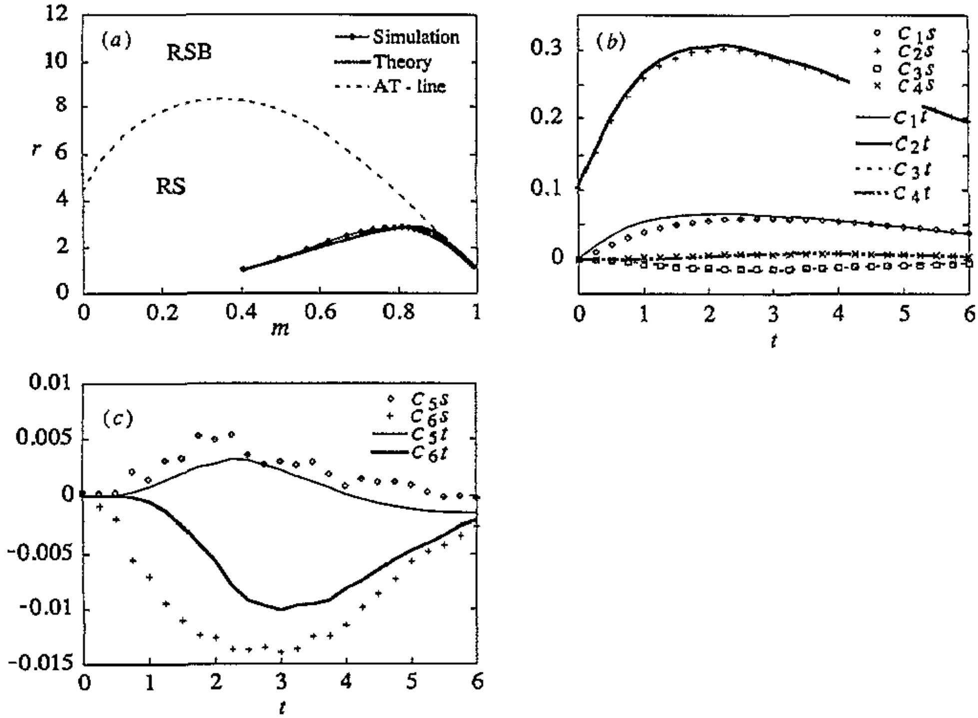

Figure 3. (a) Time-dependent behaviour of m, r (successful retrieval); $T = 0, \alpha = 0.1, N = 30\,000$ and $m_0 = 0.4$ (just above the critical value for successful retrieval). Only one example was tried in this case. Notice that r increases at an early stage and then decreases to 1. The full curve represents the CS prediction for the time evolution of the order parameters. (b) Time evolution of cumulants of noise distribution to fourth order; conditions as in (a). C_1t denotes the first cumulant of the theoretical noise distribution and C_1s the simulation result. Deviation of the theory from simulation develops at the first stage but gradually vanishes. (c) Time evolution of the fifth and sixth cumulants of noise distribution; conditions as in (a). Deviation of theory from simulation develops at the first stage but gradually vanishes.

N the trajectories of m and r do not depend on the pattern realization. In figures 2(c) and (d) we compare the cumulants of the noise distribution obtained by the replica calculation $D^{\mathrm{RS}}_{m,r}[z]$ with numerical simulations ($N = 5000, 10\,000$ and $30\,000$). It is noticed that there exist clear differences between the theory and simulations. Even when the network is in the RS region where CS conjectured exactness of their theory ($t \leqslant 2$ iteration per neuron in figure), the theory does not describe the noise distribution in a quantitative sense. As for the first and third cumulants C_1 and C_3, although it is difficult to notice the difference from the figure, there actually exist more than 30% differences between the theory and simulations. The same behaviour was obtained in the case of $\alpha = 0.2$ and $m_0 = 0.1$ where retrieval fails. These results lead us to the conclusion that the CS theory is not exact when retrieval fails, although it is a fairly good approximation, as seen in figures 1(b)–(d).

Another case is shown in figure 3 ($N = 30\,000, \alpha = 0.1$ and $m_0 = 0.4$). In this case the initial overlap m_0 is very close to, but just above, the critical value for successful retrieval. Although the effect of non-retrieval patterns r increases at an early stage, the first pattern defeats other patterns and finally retrieval succeeds. An interesting behaviour can be seen in figures 3(b) and (c). When r increases, the difference between theoretical and simulation noise distributions increases. However, the theory becomes very close to the simulation

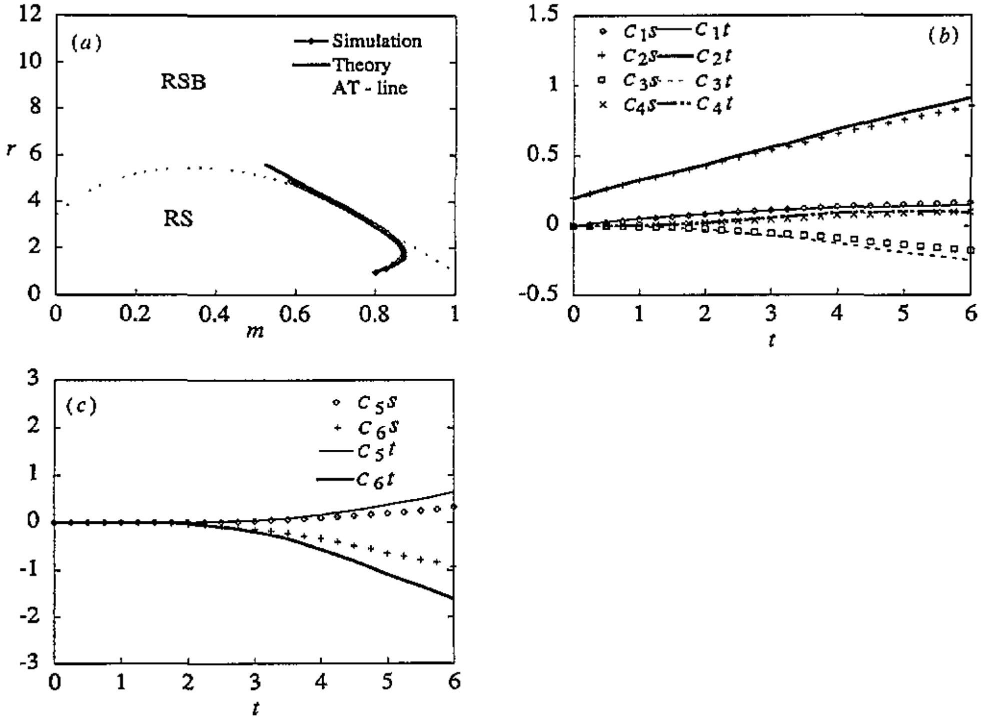

Figure 4. (*a*) Time-dependent behaviour of m, r (non-successful retrieval); $T = 0, \alpha = 0.2$, $N = 30\,000$ and $m_0 = 0.8$. Only one example was tried. Although m_0 is high, retrieval fails because $\alpha = 0.2$ is large. The full curve represents the CS prediction for the time evolution of the order parameters. (*b*) Time evolution of cumulants of noise distribution to fourth order; conditions as in (*a*). When r is close to 1, the theory stays very close to the simulations, although diverges from the simulations as soon as r becomes larger. (*c*) Time evolution of the fifth and sixth cumulants of noise distribution; conditions as in (*a*). When r is close to 1, the theory stays very close to simulations, but diverges as soon as r becomes larger.

when r becomes smaller later. Because the network stays in the RS region, we cannot explain the deviation in the early stage by ascribing it to replica-symmetry breaking.

We also tried the case of $m_0 = 0.8$ with $\alpha = 0.1$. The overlap m develops towards 1, whereas r does not increase nor decrease and stays in the very close vicinity of the initial state $r = 1$. As all the cumulants, except the second, stay zero in the time development, we omitted the figure. But the theory described the simulation well with a discrepancy of less than 1%. The same behaviour has been observed at finite temperatures where the dynamics are governed by the transition rate (1).

From the above results, we conclude that the theory describes the noise distribution quite faithfully only when the effect of other patterns is small (or r is close to 1). Figure 4 supports this conclusion. Although m_0 is high, retrieval fails because $\alpha = 0.2$ is large (overloading) in this case. When r is small, the theory stays very close to simulations. However, the theory starts to go away from simulations when r increases. As the network evolves towards equilibrium states in figures 2 and 4, the difference between the theory and simulation increases. We conclude that such a phenomenon does not result from the fact that the network state goes into the RSB region but comes from some other reasons. Firstly, although equipartitioning of probability of the network states in the subshells of m

and r is correct in equilibrium, it is not readily acceptable in dynamical, non-equilibrium processes. Moreover, there must be some limits in the idea that the network state can be described only in terms of two parameters m and r, especially when the shape of distribution becomes complicated in non-retrieval cases. According to the theory of CS, once α, m and r are given, the noise distribution is determined independently of other conditions such as the temperature and history through which the network passed during time development. We need more than two parameters to describe the dynamical processes quantitatively when retrieval fails, as suggested by Gardner *et al* [4].

As our simulation clearly showed, the theory of CS is a fairly good approximation for qualitative aspects of memory retrieval, although there definitely exist quantitative deviations from simulation results. Their method may be applicable to many situations beyond the scope of equilibrium statistical mechanics because CS did not actually use the specific form of the interaction (2) in their derivation of the noise distribution.

Acknowledgments

One of the authors (HN) thanks Dr Tom Coolen and Professor David Sherrington for discussions. Use of the fast random generator RNDTIK, written by N Ito and Y Kanada, is gratefully acknowledged.

References

[1] Coolen A C C and Sherrington D 1994 *Phys. Rev.* E **49** 1921

[2] Coolen A C C and Sherrington D 1993 *Phys. Rev. Lett.* **71** 3886

[3] Nishimori H and Ozeki T 1993 *J. Phys. A: Math. Gen.* **26** 859

[4] Gardner E, Derrida B and Mottishaw B 1987 *J. Physique* **48** 741

[5] Amari S and Maginu K 1988 *Neural Networks* **1** 63

[6] Rieger H, Schreckenberg M and Zittartz J 1988 *Z. Phys.* B **72** 523

[7] Zagrebnov V A and Chvyrov A S 1989 *Sov. Phys.–JETP* **68** 153

[8] Patrick A E and Zagrebnov V A 1991 *J. Phys. A: Math. Gen.* **24** 3413

[9] van Hemmen J L and Kühn R 1991 *Models of Neural Networks* ed E Domany, J L van Hemmen and K Schulten (Berlin: Springer) pp 1–105