Retrieval dynamics of associative memory of the Hopfield type

1993 J. Phys. A: Math. Gen. 26 859

(http://iopscience.iop.org/0305-4470/26/4/013)

J. Phys. A: Math. Gen. 26 (1993) 859–871. Printed in the UK

Retrieval dynamics of associative memory of the Hopfield type

Hidetoshi Nishimori and Tomoko Ozeki

Department of Physics, Tokyo Institute of Technology, Oh-okayama, Meguro-ku, Tokyo 152, Japan

Received 11 May 1992, in final form 28 September 1992

Abstract. We have investigated the dynamics of retrieval processes of an associative memory of the Hopfield type. For synchronous dynamics, we have generalized the theory of Amari and Maginu, which enables us to treat the intermediate processes of memory retrieval in terms of a few simple macrovariables, to the finite-temperature case. The resulting phase diagram in the equilibrium limit agrees qualitatively well with that from equilibrium statistical mechanics. We have carried out Monte Carlo simulations to clarify the limit of applicability of their theory. We have found that their basic assumption, an independent Gaussian distribution of the noise term with a time-dependent variance, is satisfied if the network succeeds in retrieval. When retrieval fails, the distribution of noise is non-Gaussian from very early stages of time development. For asynchronous dynamics, we propose a time-dependent Ginzburg–Landau (TDGL) approach, which simply expresses a downhill motion of the network in the free energy landscape. The resulting flow diagram in a phase space describes the behaviour of the network when it is close to equilibrium.

1. Introduction

Statistical mechanics has been a powerful tool to investigate equilibrium properties of neural networks of the Hopfield type [1]. By exploiting the analogy with the Ising model, Amit *et al* [2,3] successfully derived the free energy and equations of state to describe various macroscopic phases in the presence of thermal noise. The same method has been applied to solve the case of synchronous updating processes [2,4]. As has been shown by Peretto [5], synchronous dynamics applied to the Hopfield-type network yields an equilibrium state corresponding to a slightly modified effective Hamiltonian. This fact enables one to apply the techniques of equilibrium statistical mechanics to the synchronous network.

Dynamics of retrieval processes is more difficult to treat than equilibrium properties mainly because no general prescription corresponding to the Boltzmann–Gibbs framework is available. One therefore has to work out a method appropriate for each problem. The significance of investigation of dynamics is never negligible in spite of such inherent difficulties. The dependence of the final equilibrium state on the initial condition can be clarified through dynamics. This means that we can learn the size and shape of the basin of attraction. Non-trivial behaviour of the network during the retrieval process, as will be shown later, is also clarified only by a direct dynamical treatment.

For these and other reasons, several authors have proposed various ways of description of dynamics. Meir and Domany [6] used a layer-structured network in

0305-4470/93/040859+13$07.50

which embedded patterns differ from layer to layer. For such a case, they showed that it is possible to derive a closed set of evolution equations of macrovariables. An asymmetric diluted network is another instance for which an exact analysis can be carried out [7]. In a more general case of the usual symmetric network with an extensive number of patterns embedded in it, it is necessary to introduce an infinite number of parameters to describe time evolution without approximations [8,9]. Riedel *et al* [10] used an effective equilibrium Hamiltonian, related to the equilibrium free energy, to derive a compact set of differential equations. Horner *et al* [11] instead treated dynamics directly in terms of the generating functional formalism. Under certain assumptions on the asymptotic form of correlation functions, they were able to reproduce the correct behaviour of the network in the initial and final stages of time development.

Amari and Maginu [12] employed a signal-to-noise ratio analysis for a synchronous network. Under the assumption of independent Gaussian distribution of the noise contribution, they discussed the time dependence of macrovariables and the shape of the basin of attraction. Other investigations along similar lines have subsequently appeared [13,14]. An advantage of this type of theory is that the basic assumption is simple and clear so that one can verify it in rather straightforward numerical simulations as shown below. A purpose of the present paper is to check the validity of the Amari–Maginu theory and generalize it to a finite-temperature case. We also propose a simple set of equations, the TDGL equation, to discuss the asynchronous dynamics.

In section 2, we discuss the properties of a synchronous network with extensively many random patterns embedded by the Hebb rule. We generalize the Amari–Maginu approach to the finite temperature problem. A closed set of recursion equations are presented to describe the macroscopic time development of the network in the process of memory retrieval. In particular, the behaviour in the equilibrium limit is compared with the predictions of equilibrium statistical mechanics. We then show the results of our Monte Carlo simulations on the network with the aim to clarify under what conditions the assumption of Amari and Maginu of a Gaussian distribution of noise is satisfied. We turn to the asynchronous dynamics in section 3. When the network state is close to equilibrium, its time development may be described as a downhill motion in the free energy landscape. This idea can be realized as the TDGL (time-dependent Ginzburg–Landau) equation; we assume that the time derivative of a macrovariable is proportional to the derivative of the free-energy functional. The TDGL equation derived from the free energy functional by Amit *et al* [3] turns out to describe well the final stages of memory retrieval in the asynchronous network. The results of this approach are compared with those in previous sections for the synchronous network. The final section is devoted to discussions.

2. Retrieval processes for synchronous dynamics

2.1. Recursion relations

Let us consider a network of two-state neurons $\{S_i^t = \pm 1\}_{i=1,..N}$ with a synchronous stochastic updating dynamics

$$\mathrm{Prob}(S_i^{t+1}) = \tfrac{1}{2}(1 + S_i^{t+1} \tanh \beta h_i^t) \tag{1}$$

where h_i^t denotes the input to the ith neuron at time t:

$$h_i^t = \sum_{j \neq i} J_{ij} S_j^t. \tag{2}$$

We embed P random patterns $\{\xi_i^\mu = \pm 1\}$ ($\mu = 1, \cdots, P$) according to the rule

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu. \tag{3}$$

By the prescription (3), each of the embedded patterns is an approximate equilibrium state of the dynamics (1) if the temperature $T = 1/\beta$ is sufficiently low [1–4].

It is convenient to introduce a macrovariable, the overlap of the current network state with the μth pattern,

$$m_t^\mu = \frac{1}{N} \sum_i \xi_i^\mu S_i^t. \tag{4}$$

Finiteness of m_t^μ in the infinite time limit would indicate a successful retrieval of the μth pattern. Our task is thus to derive a recursion relation of m_t^μ (and another related macrovariable) to describe the macroscopic time development of the network from an initial state. Hereafter we restrict ourselves to the case in which only one of the embedded patterns is retrieved, which we choose as the first pattern ($\mu = 1$). Let us denote m_t^1 simply as m_t. Without loss of generality, we can set

$$\xi^1 = (1, 1, \ldots, 1). \tag{5}$$

The input (2) can then be divided into two parts, the signal and noise,

$$h_i^t = \sum_j J_{ij} S_j^t = \frac{1}{N} \sum_j \sum_\mu \xi_i^\mu \xi_j^\mu S_j^t = m_t + N_i^t \tag{6}$$

where the first term in the final expression comes from the $\mu = 1$ contribution (signal for retrieval of the first pattern) in the intermediate summation, and the second term is the noise

$$N_i^t = \frac{1}{N} \sum_{\mu \geqslant 2} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu S_j^t. \tag{7}$$

The noise N_i^t represents disturbing effects from non-retrieved patterns ($\mu \geqslant 2$).

To proceed further, Amari and Maginu [12] regarded the noise N_i^t as a random variable and assumed that it obeys a Gaussian distribution with a vanishing mean and a time-dependent variance σ_t^2. If terms in the summation on the right-hand side (RHS) of (7) were independent random variables (which is true for $t = 0$), the central limit theorem would apply to lead to a Gaussian distribution of N_i^t in the limit of large N. In general such a situation is not realized, however, since S_j^t depends on the set $\{\xi_i^\mu\}$ because of the updating processes (1) in previous time steps. Amari and Maginu nevertheless assumed a Gaussian distribution and investigated the consequences. As we will see later, this assumption is approximately valid under certain conditions. We *stress here that, even if the Gaussian assumption turns out to be valid, it is not* because the central limit theorem applies but for some other unknown reasons. It is also important to note that they assumed that N_i^t obeys a Gaussian distribution independently at each neuron i at a given time step t.

The macroscopic overlap m_{t+1} is now expressed by

$$m_{t+1} = \frac{1}{N}\sum_i \mathrm{E}[S_i^{t+1}] = \frac{1}{N}\sum_i \tanh\beta(m_t + \sigma_t\epsilon_i) \tag{8}$$

where ϵ_i is a normalized Gaussian variable defined independently at each i. We have taken the expectation value, symbolized by $\mathrm{E}[\ldots]$, with respect to the stochasticity represented by (1). The last expression in (8) converges to the ensemble average of $\tanh\beta(m_t + \sigma_t\epsilon_i)$ with respect to the Gaussian distribution of ϵ_i in the large N limit to yield

$$m_{t+1} = \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \tanh\beta(m_t + \sigma_t z). \tag{9}$$

It is necessary to derive a recursion relation of σ_t, which is another macrovariable, to describe the strength of disturbance from non-retrieved patterns. The recursion relation of the variance can be calculated by evaluating the square of the noise N_i^t and taking account of the dependence of the current state S_j^t on the embedded patterns. This last dependence can actually be calculated by analysing the input in the preceding time step. The only point to be modified from calculations by Amari and Maginu [12] is that one replaces the function $\mathrm{sign}(h_i^t)$ by $\tanh\beta h_i^t$ reflecting finiteness of temperature. The result reads

$$\sigma_{t+1}^2 = \alpha + 2\alpha\beta m_{t+1} m_t h(\beta m_t, \beta\sigma_t) + \beta^2\sigma_t^2 h(\beta m_t, \beta\sigma_t)^2 \tag{10}$$

where

$$h(\beta m_t, \beta\sigma_t) = \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \operatorname{sech}^2\beta(m_t + \sigma_t z). \tag{11}$$

The parameter α is defined as $\alpha = P/N$.

The recursion relations (9) and (10) determine the macroscopic dynamical behaviour of the network. Numerical solutions of these equations reveal that, quite similarly to the zero-temperature case [12], the network successfully retrieves the first embedded pattern if α is small and the initial overlap is larger than a critical value $m_0(\alpha)$. If α is large, any initial condition leads to a non-retrieval state $m_t \to 0$ at the final stages $t \to \infty$. It turns out that a successful retrieval ($m_\infty > 0$) is possible for α smaller than $\alpha_c(T)$. The final overlap m_∞ drops to zero discontinuously at α_c. These properties in the infinite time limit are in agreement with those derived from the equilibrium statistical mechanics [3,4].

As shown in figure 1, α_c decreases continuously from 0.1597 at $T = 0$ and vanishes at $T = 1$. The critical capacity calculated by equilibrium statistical mechanics is also drawn in figure 1 for synchronous [4] as well as asynchronous [3] dynamics. One should note the qualitative resemblance of the three curves in figure 1. Our method is simple enough to use only the Gaussian assumption of the noise term N_i^t, and yet reproduces well the results from quite a different method, namely, equilibrium statistical mechanics. We stress qualitative similarity of the critical curves in figure 1, leaving the quantitative difference as a minor problem. We do not intend to claim any mathematical rigor of the final expressions (9) and (10); these equations are useful approximations for the purpose of clarifying the dynamics of retrieval processes which cannot be treated by equilibrium statistical mechanics, especially about the early stages of time development. Some examples of such transient phenomena will be given in the next subsection. Close to the critical capacity $\alpha_c(T)$, the assumption of Gaussian noise distribution may not work very well, which may be one of the reasons for the quantitative difference between our results and those of statistical mechanical calculations.

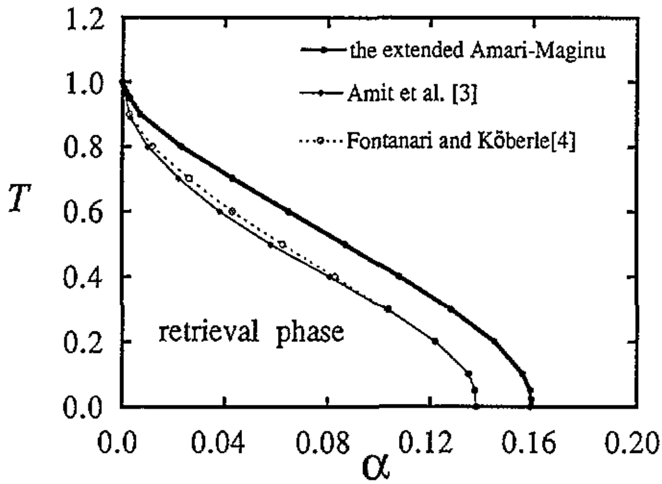

Figure 1. Equilibrium phase diagram obtained from the generalized Amari–Maginu theory. The critical capacity α_c decreases continuously from 0.1597 at $T = 0$ value and vanishes at $T = 1$. For reference, the critical capacity calculated by equilibrium statistical mechanics is also given for synchronous [4] as well as asynchronous [3] dynamics.

We remark here that (9) and (10) do not have a paramagnetic solution $m_t = 0$ and $\sigma_t = 0$, while the statistical mechanical approach yields such a state at high temperatures. This is understandable because, as will be discussed in the next subsection, the present method is likely to fail in the case of unsuccessful retrieval: we do not expect the Amari–Maginu method to predict the behaviour of the network outside the successful retrieval region.

To see some of the dynamical aspects of recursion relations (9) and (10), let us take the limit of low storage $\alpha \to 0$. In such a limit, an equilibrium solution of the recursion relation is $\sigma = 0$ and $m = \tanh(\beta m)$, which agrees with the prediction of statistical mechanics [2]. Thus the effects of non-retrieved patterns, which are represented by the variance σ, vanish in the low-storage limit. In the low-temperature and small-α limit, the recursion relation reduces to

$$\sigma_t = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{m_{t-1}^2}{2\sigma_{t-1}^2}\right).$$

This equation indicates a very quick approach to the noiseless ($\sigma = 0$) solution.

Although the high-temperature region is outside the range of applicability of the present method, it may be of some interest to see what type of behaviour is derived within the Amari–Maginu framework. The recursion relations (9) and (10) in the high temperature limit $\beta \to 0$ are expressed as

$$m_t = \beta m_{t-1} \qquad \sigma_t^2 = \alpha + \beta^2 \sigma_{t-1}^2.$$

Approach to the equilibrium value $m = 0, \sigma^2 = \alpha/(1-\beta^2)$ is written as

$$m_t = \mathrm{e}^{-t/\tau} m_0 \qquad \sigma_t = \sigma + \mathrm{e}^{-2t/\tau} S_0$$

where m_0 and S_0 are determined by the initial condition and $\tau = 1/\log T$. As T approaches 1, the relaxation time τ is seen to diverge, a critical slowing down. The actual critical temperature depends upon α and is smaller than 1; the divergence of τ at $T = 1$ would be a consequence of the high temperature approximation.

2.2. *Simulations*

We have carried out Monte Carlo simulations to check the validity of the basic assumption of the Amari–Maginu theory, namely, an independent Gaussian distribution of the noise term N_i^t at each i. The number of neurons N was 5000, 7000 and 9000. The parameter α was chosen to be 0.08 and 0.20 for each N. We have tried $m_0 = 0.1, 0.2, \ldots, 0.9$ as the initial condition. The temperature was set to zero.

To see the distribution of the noise term, we have separated numerically the input h_i^t to the true signal m_t and the rest N_i^t as in (6). Strictly speaking, randomness or stochasticity of N_i^t comes from the randomness of $\{\xi_i^\mu\}$. This means that we should generate many sets of random patterns and see the stochastic distribution of N_i^t at a fixed i. However, generating very many sets of random patterns $\{\xi_i^\mu\}$, of the order of 10^4 sets for good statistics, and performing simulations for all of them require inhibitingly large computing efforts. We have instead taken statistics of N_i^t by varying i from 1 to N at each time step t for a fixed set of random patterns $\{\xi_i^\mu\}$. It is not obvious *a priori* that these two procedures amount to the investigation of the same quantities. However, the equivalence of the configurational (randomness) average and the spatial average in spin glasses [15] is an encouraging fact to justify of our simplified method of statistics. We have also checked by investigating several different sets of random patterns that the following results do not depend in any remarkable way on the choice of the set of random patterns.

To characterize the noise distribution quantitatively, we have calculated cumulants of the distribution to fourth order. Fifth- and higher-order cumulants are difficult to calculate with satisfactory accuracy because such a calculation requires subtractions of high-order moments, which are large numbers, leading to larger statistical errors by losing significant digits. For this reason, our criterion is that a distribution is Gaussian if the cumulants of third and fourth orders are vanishing within statistical errors determined by the usual $1/\sqrt{N}$ rule.

Some of the results are shown in figure 2. Figure 2(a) represents the case $N = 9000$, $\alpha = 0.08$ and $m_0 = 0.5$. The overlap m_t reaches a saturation close to unity after several time steps, a successful retrieval. The actual time dependence of the overlap was in excellent agreement with the Amari–Maginu prediction (9)

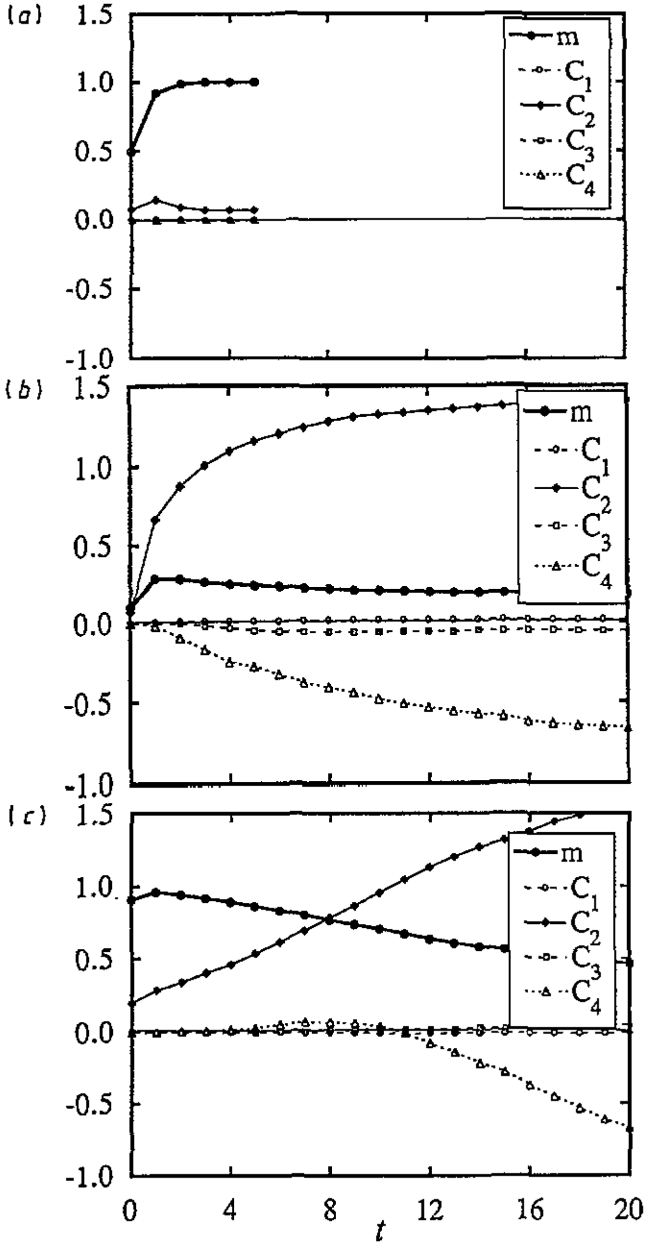

Figure 2. Time dependence of cumulants of the noise. T=0 and N=9 000. C_i $(i = 1,\ldots,4)$ represents the ith cumulant of the noise term. (a) $\alpha = 0.08, m_0 = 0.5$ (b) $\alpha = 0.08, m_0 = 0.1$ (c) $\alpha = 0.2, m_0 = 0.9$ It is seen that only when the memorized pattern is successfully retrieved as in (a), the third and fourth-order cumulants stay in very small values, i.e. the Gaussian assumption is satisfied.

and (10). The first-order cumulant vanishes at all time steps while the second-order cumulant, the variance σ_t^2, followed well the prediction of (9) and (10) and reached a steady state as soon as m_t did. The third and fourth-order cumulants stayed very small, effectively vanishing within statistical errors, at all t. These observations were shared by the smaller N cases, 7000 and 5000, with the same α and m_0. Figure 2(b) is for $N = 9000$, α=0.08, $m_0 = 0.1$. Retrieval apparently fails under this condition: the initial overlap m_0 is too small to warrant retrieval. The first- and third-order cumulants stay in a close vicinity of zero, while the second- and fourth-order cumulants grow rapidly from very early stages of time development. It should also be added that significant size dependence of m_t and σ_t was observed in this case, and the prediction (9) and (10) was not followed well quantitatively by numerical data. Similar behaviour was found for the case $N = 9000$, $\alpha = 0.20$, $m_0 = 0.8$ as in figure 2(*c*).

Our conclusion obtained from checking data for all parameter combinations is that the distribution of noise is very close to be Gaussian when memory retrieval is successful. It is remarkable that the distribution is likely to be Gaussian at all time steps for successful retrieval cases. The noise 'anticipates' a successful retrieval from the outset of time development. We have also checked independence of noise N_i^t at each i by dividing the whole network into two parts with the same size, A and B, and looking at the covariance of N_i^t between $i \in$ A and $i \in$ B. To be more precise, we calculated the covariance $\mathrm{Cov}(N_i^t, N_{i+N/2}^t)$ by running i from 1 to $N/2$ for simplified statistics as before. The result indicated a vanishing covariance for any values of the parameters within statistical errors. We therefore conclude that the Amari–Maginu theory describes the dynamics of the network quite faithfully when the memory retrieval is successful. This conclusion is supported also by preliminary data of simulations for finite-temperature cases.

Another interesting behavioural feature of the network was found in the case of large values of m_0, 0.9 for instance, and $\alpha = 0.20$. Although the memory retrieval fails for such a large α, and correspondingly the third- and fourth-order cumulants grow at final stages of time development, those two cumulants remain close to zero until a certain critical time step t_c is reached; this is seen in figure 2(*c*). Beyond the critical time, the cumulants, in particular the fourth cumulant, increase rapidly. The critical time step, typically of the order of $t_c = 10$, depends rather strongly upon the sample of randomness $\{\xi_i^\mu\}$ and system size N. We could not see any clear systematic tendency, suggested by Amari and Maginu [12], of $t_c(N)$ going to infinity as N increases. Much larger N and more extensive analysis of sample dependence are required to settle this issue.

It is not surprising that the simple Gaussian assumption fails if memory is not retrieved; the network would fall into a spin-glass state in such a case, which is likely to require a much more sophisticated treatment than the two-parameter theory.

Before closing this section, let us remark that we do not claim to have confirmed the rigorousness of the Amari–Maginu theory in the case of successful retrieval. Our point is that the third- and fourth-order cumulants do not grow with time and stay very small if not strictly vanishing; it is impossible to prove a vanishing value by numerical simulations. Very small values, which are uniform (non-growing) in time, would be sufficient for the validity of regarding the Amari–Maginu theory as a good approximation.

3. TDGL approach

A downhill motion of the network state in the free-energy landscape is an intuitively appealing idea. The time-dependent Ginzburg–Landau (TDGL) equation is a phenomenological equation to express the downhill motion. It is in general written for a macrovariable m as

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{1}{t_0}\frac{\partial f}{\partial m} \tag{12}$$

where t_0 sets the time unit and f denotes the free energy functional written in terms of m. If we use the free energy functional calculated within the framework of equilibrium statistical mechanics on the RHS of (12), the resulting equation will describe the behaviour of the network near equilibrium. Actually, in the case of the infinite-range ferromagnetic Ising model, this type of approach is known to yield the exact evolution equation [16].

The free energy functional for the asynchronous network has been given by Amit *et al* [3] under the assumption of replica symmetry. For a single-pattern retrieval, it reads

$$f = \frac{1}{2}\alpha + \frac{1}{2}m^2 + \frac{\alpha}{2\beta}\left(\ln(1-\beta+\beta q) - \frac{\beta q}{1-\beta+\beta q}\right) + \frac{\alpha\beta r(1-q)}{2} - \beta^{-1}\int \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \ln 2\cosh\beta(\sqrt{\alpha r}z + m). \tag{13}$$

Three order parameters appear in (13). The overlap m has the same meaning as in (4). The spin glass order parameter q denotes the degree of random freezing of the state of each neuron. The last parameter r represents the accidental overlap of the network state with non-retrieved patterns

$$r = \frac{1}{\alpha}\sum_{\mu=2}^{P}(m^\mu)^2$$

which bears a physical meaning similar to that of the variance σ_t^2 of noise from non-retrieved patterns used in the previous section. The TDGL equation for the free energy (13) is now

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{1}{t_0}\frac{\partial f}{\partial m} = -\frac{1}{t_0}\left\{m - \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\tanh\beta(m+\sqrt{\alpha r}z)\right\} \tag{14a}$$

$$\frac{\mathrm{d}r}{\mathrm{d}t} = -\frac{1}{t_0}\frac{\partial f}{\partial r} = -\frac{\alpha\beta}{2t_0}\left\{-q + \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\tanh^2\beta(m+\sqrt{\alpha r}z)\right\} \tag{14b}$$

$$\frac{\mathrm{d}q}{\mathrm{d}t} = -\frac{1}{t_0}\frac{\partial f}{\partial q} = -\frac{\alpha\beta}{2t_0}\left\{\frac{q}{(1-\beta+\beta q)^2} - r\right\}. \tag{14c}$$

We have assumed here that the time scale t_0 is common to all three equations, which does not affect the final qualitative behaviour of the network.

As in the theory of spin glasses [15], a steady-state solution of (14) is not in general a local minimum of the equilibrium free energy (13). As shown in figure 3, the free energy is at minimum for perturbations in m, but this is not the case for q and/or r. This difficulty seems to be originated in the replica trick used in the derivation of (13). Expressions involving a summation over two different replica indices, such as $\sum_{\rho\neq\delta} r_{\rho\delta}q_{\rho\delta}$, acquire a factor $n(n-1)$ by the replica symmetry assumption $r_{\rho\delta} = r$, $q_{\rho\delta} = q$, where n denotes the number of replicas. In the limit of $n \to 0$, the factor $n(n-1)$ becomes negative, which would result in the inverted landscape of the free energy.

Thus it would not to lead to meaningless conclusions if we invert the free energy along the direction of a negative eigenvalue of the curvature matrix

$$M = \begin{pmatrix} \partial^2 f/\partial r^2 & \partial^2 f/\partial r\partial q \\ \partial^2 f/\partial r\partial q & \partial^2 f/\partial q^2 \end{pmatrix} \tag{15}$$

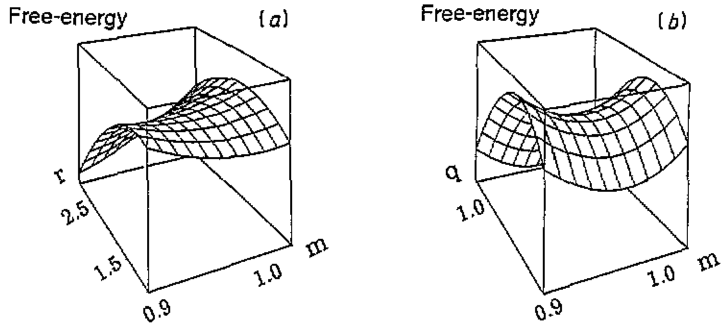

Figure 3. The free energy landscape; (*a*) The variable q is fixed at q_e (equilibrium value) and the free energy is depicted as a function of m and r. (*b*) The variable r is fixed at r_c (equilibrium value). We can see that a steady-state solution of (20) is not a local minimum of the free energy.

evaluated at the saddle point. Note that the equation for m is left intact since the free energy is always minimum with respect to this variable. For any combinations of the parameters α and T, we have found that the matrix M has one positive and one negative eigenvalues. Let us choose the x axis around a saddle point as the direction of the eigenstate corresponding to the positive eigenvalue and y axis as that of the negative eigenvalue. We then use the modified TDGL equation

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{\partial f(m,x,y)}{\partial m} \tag{16a}$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\partial f(m,x,y)}{\partial x} \tag{16b}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = +\frac{\partial f(m,x,y)}{\partial y}. \tag{16c}$$

We have set $t_0 = 1$ for notational simplicity. Numerical solutions of (16) are transformed back to the representation in terms of r and q.

We solved (16) for various initial conditions. The initial value of the overlap m_0 can be set arbitrarily between 0 and 1. There is no obvious choice of q_0, and we tried a variety of q_0 to find that the results are hardly dependent upon this initial condition. In particular, $q(t)$ converges very quickly, as compared to $m(t)$ and $r(t)$, to a common equilibrium value starting from a wide range of q_0. This may imply that essentially only two macrovariables are sufficient to describe the dynamics of retrieval processes as was the case in the Amari–Maginu theory for synchronous dynamics. The initial value for r is uniquely determined from the condition that the initial network state has a finite overlap only with the pattern to be retrieved. That is, since

$$\langle (m_0^\mu)^2 \rangle = \frac{1}{N^2}\Big\langle \sum_i \sum_j \xi_i^\mu \xi_j^\mu S_i^0 S_j^0 \Big\rangle = \frac{1}{N}$$

then for $\mu \geqslant 2$, we have

$$r_0 = \frac{1}{\alpha}\sum_{\mu \geqslant 1} \langle (m_0^\mu)^2 \rangle \simeq \frac{1}{\alpha} P \frac{1}{N} = 1.$$

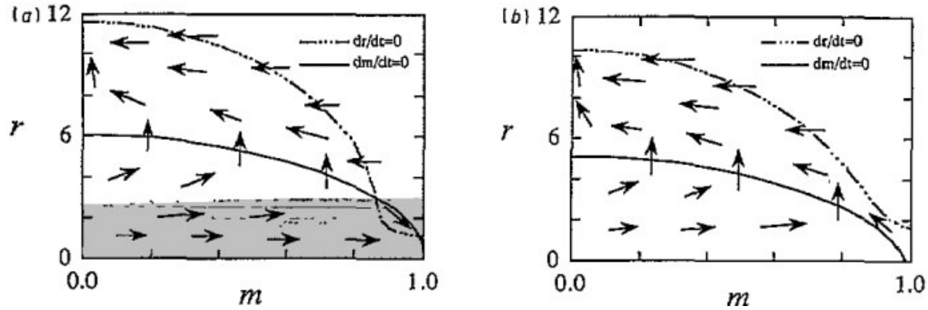

Figure 4. Time evolution of the network in the m versus r diagram; (a) $\alpha = 0.1$, $T = 0.3$ (retrieval phase). There are two fixed-point attractors, one near $m = 1$ and another at $m = 0$. The former is a successful retrieval fixed point while the latter represents failure. The basin of attraction of the retrieval fixed point is shown shaded. (b) $\alpha = 0.1$, $T = 0.4$ (spin-glass phase). The only fixed point is at $m = 0$. The whole phase space is in the basin of attraction of this spin-glass fixed point.

We actually tried other values of r_0 because the flow diagram thus written enables us to understand the dynamics in a transparent way as shown below.

Figure 4(a) displays the time development of the network by a set of arrows in the m versus r phase diagram for the case of $\alpha = 0.1$ and $T = 0.3$. The spin glass order parameter q is fixed to an equilibrium constant for the reason mentioned above. Along the two solid lines, time derivative of one of the order parameters is zero, which means that the flow is either vertical or horizontal on the lines. There are two fixed-point attractors, one near $m = 1$ and another at $m = 0$. The former is a successful-retrieval fixed point while the latter represents failure. The basin of attraction of each fixed point is separated by a definite curve below which all points are attracted to the successful-retrieval fixed point. (We thank Nakamura for suggesting this type of analysis.) Apparently all physically realizable initial states, which have $r_0 = 1$, are attracted to the successful retrieval fixed point irrespective of m_0. Simulations rather suggest that there is a finite critical m_0 below which retrieval fails [11] as was the case for synchronous dynamics discussed in previous sections. We conclude therefore that the TDGL approach using the equilibrium free energy does not describe the initial condition dependence of the network behaviour in a convincing manner. This fact was somewhat expected because the free energy functional calculated within the framework of equilibrium statistical mechanics is not likely to dominate the network properties in the early stages of time development starting from an arbitrary non-equilibrium state. Only the final stages can be treated by the present approach.

When the parameter is in the region corresponding to the spin glass phase, the flow diagram looks like figure 4(b) in which $\alpha = 0.1$ and $T = 0.4$. There is no crossing of the two curves $\mathrm{d}m/\mathrm{d}t = 0, \mathrm{d}r/\mathrm{d}t = 0$, so that the only fixed point is at $m = 0$. The whole phase space is in the basin of attraction of this spin-glass fixed point.

4. Discussion

We have investigated the retrieval dynamics of the network of two-state neurons connected by the efficacy (3). For synchronous dynamics, the assumption of independent Gaussian distribution of the noise term N_i^t is pointed out to have a crucial significance in clarifying the limit of applicability of the Amari–Maginu method. An advantage of this approach is that the basic assumption is clearly stated and hence is amenable to direct numerical verifications. Our Monte Carlo simulations reveal that this assumption is likely to hold, at least to a quite good approximation, if memory retrieval succeeds in the final stages of time development. The noise distribution deviates from Gaussian from very early stages if retrieval eventually fails. Therefore, at early and intermediate stages, the noise contains more information on the final result than the most important macrovariable, the overlap m_t.

Of course, the noise distribution has much more degrees of freedom than the single macrovariable m_t, and one may be inclined to attribute this difference in the degrees of freedom to the difference in the amount of information. A real surprise is, however, that the noise has actually only one degree of freedom, the variance σ_t^2, in the case of successful retrieval. This fact has been found numerically. We do not claim that this is always the case in a strict sense; higher-order cumulants may be very small but non-vanishing. An important point is that the higher-order cumulants are observed to stay very small uniformly in time. This uniformity in time enables us to treat the dynamics to a very good approximation within the framework of the present theory.

Amari and Maginu [12] tried an improvement of their theory by allowing a non-vanishing mean of the Gaussian noise distribution. The result, however, was not impressive because, first, the final state became independent of the initial condition m_0 in the whole range $0 < m_0 < 1$, and second, the equilibrium value m showed continuous dependence on α. Such behaviour is in conflict with simulations as shown in section 2. It is a future problem to clarify why such an apparent improvement leads to less reliable results. A hint may be that appropriateness of approximation is not necessarily a monotonic function of the degree of approximation, as is often the case for asymptotic expansions. We note in passing that the Amari–Maginu expressions for the macrovariables $m_{t=2}$, $b_{t=1}$ and $\sigma_{t=1}$, where b_t is the mean value of the noise, agree with the exact formula for the first few steps derived in [8].

We generalized the Amari–Maginu theory to the finite-temperature problem. The result on the equilibrium phase diagram agrees qualitatively well with that from statistical mechanical calculations. The slight quantitative difference may be attributed to a degraded Gaussian approximation near critical capacity. The present formulation is quite simple in its principle and yet yields fruitful information on dynamical processes as well as on equilibrium properties. Even in comparison with other approaches to dynamics of retrieval processes [6–11, 13, 14], the present method is unique in its simplicity and generality. Only the assumption of Gaussian noise distribution is sufficient to derive all necessary equations which describe the network behaviour in a reasonable manner.

The network properties near equilibrium should be characterizable by the downhill motion in the free energy landscape. The phenomenological TDGL equation represents such a situation. The flow diagram obtained from TDGL equation (figure 4) reveals global motions of the network near equilibrium.

Acknowledgments

We thank Tota Nakamura for discussions and useful suggestions. Use of the fast random number generator RNDTIK written by N Ito and Y Kanada is gratefully acknowledged. This work was supported by the Grant-in-Aid for Scientific Research on Priority Areas by the Ministry of Education, Science and Culture of Japan.

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
[2] Amit D J, Gutfreunt H and Sompolinsky H 1985 *Phys. Rev.* A **32** 1007
[3] Amit D J, Gutfreunt H and Sompolinsky H 1987 *Ann. Phys.* **173** 30
[4] Fontanari J F and Köberle R 1988 *J. Physique* **49** 2475
[5] Peretto P 1984 *Biol. Cybern.* **50** 51
[6] Meir R and Domany E 1987 *Phys. Rev. Lett.* **59** 359
[7] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167
[8] Gardner E, Derrida B and Mottishaw B 1987 *J. Physique* **48** 741
[9] Watkin T L H and Sherrington D 1991 *J. Phys. A: Math. Gen.* **24** 5427
[10] Riedel U, Kühn R and van Hemmen J L 1988 *Phys. Rev.* A **38** 1105
[11] Horner H, Bormann D, Frick M, Kinzelbach H and Schmidt A 1989 *Z. Phys.* B **76** 381
[12] Amari S and Maginu K 1988 *Neural Networks* **1** 63
[13] Zagrebnov V A and Chvyrov A S 1989 *Sov. Phys.-JETP* **68** 153
[14] Opper M, Kleinz J, Köhler H and Kinzel W, 1989 *J. Phys. A: Math. Gen.* **22** L407
[15] Binder K and Young A P 1986 *Rev. Mod. Phys.* **58** 801
[16] Amit D J 1989 *Modeling Brain Function* (Cambridge: Cambridge University Press) ch 3