Coexistence of uncorrelated and correlated attractors in a nonmonotonic neural network

1999 J. Phys. A: Math. Gen. 32 5551

(http://iopscience.iop.org/0305-4470/32/30/303)

J. Phys. A: Math. Gen. **32** (1999) 5551–5562. Printed in the UK PII: S0305-4470(99)00257-7

Coexistence of uncorrelated and correlated attractors in a nonmonotonic neural network

Tomoki Fukai†, Tomoyuki Kimoto‡, Makoto Doi† and Masato Okada§

† Department of Electronics, Tokai University, CREST, JST, Hiratsuka, Kanagawa 259-12, Japan
‡ Department of Electrical Engineering, Oita National College of Technology, 1666 Maki, Oita City 870-0152, Japan
§ Kawato Dynamic Brain Project, ERATO, JST, 2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Received 22 December 1998, in final form 26 April 1999

Abstract. A neural network with correlated attractors *à la* Griniasty *et al* (1993 Conversion of temporal correlations between stimuli to spatial correlations between attractors *Neural Comp.* **5** 1–17) is studied for a nonmonotonic response function in the cases of both finite and extensive memory loading. In the finite loading case with symmetric synaptic connections, the number of neighbouring memory patterns correlated with an attractor tends to be decreased as the degree of nonmonotonicity is increased. To study the extensive loading case, we generalize the previously proposed self-consistent signal-to-noise analysis to associative memory models with a general class of local Hebbian synaptic connections. Applying the results to the present model, we derive analytic expressions for the equilibrium order parameters. New finding includes the coexistence of uncorrelated and correlated attractors in the obtained phase diagram. This fact, which is confirmed by extensive numerical simulations, may have significant implications for long-term visual memory.

1. Introduction

Physiological experiments by Miyashita and Chang [1] reported that neurons in the anterior ventral temporal cortex of monkeys are highly selective towards a few of 100 coloured fractal patterns. Further study of the development of selectivity [2] revealed that the temporal correlations expressed by the serial position number of stimuli during learning are converted into the spatial correlations of the stored activity patterns. An attractor activity pattern has correlations with up to five neighbouring attractor patterns in the serial position number. To account for this temporo–spatial conversion, Griniasty *et al* [3] proposed a modified Hebbian learning rule that includes symmetric cross-correlation terms between uncorrelated memory patterns. Their model consisting of formal two-state neurons, as well as a more realistic version obtained later [4], succeeded in replicating the correlated attractors obtained by the Miyashita's experiments.

In this study, we examine the effects of the symmetric cross-correlation terms in synaptic connections in the case of nonmonotonic neural networks for both finite and extensive loading cases. Single neurons display monotonic increases in firing rates as input currents are increased. However, there are many sources of inhibition in biological neural networks, and the effective neural processing units comprising a small number of excitatory and inhibitory neurons possibly exhibit more complicated response profiles. Nonmonotonic response functions, which represent one of many possible response profiles, are known to

0305-4470/99/305551+12$30.00

improve the retrieval performance to a large extent [5–9]. Furthermore, the computational validity and biological credibility of nonmonotonic neurons were emphasized based on the history-dependent Bayesian neural dynamics introduced into associative memory [10, 11].

In studying the extensive loading case, we extend the previously proposed self-consistent signal-to-noise analysis (SCSNA) to associative memory models with a general class of local Hebbian learning rule. The SCSNA provides a powerful method to analyse the associative memory models. However, the method could be applied only for the classes of simple Hebbian learning rules. In particular, whether the method can be applicable for the present type of synaptic connections was not clear due to the finite correlations among attractors. The phase diagram obtained by the SCSNA reveals an interesting result, namely the coexistence of uncorrelated and correlated attractors in a certain region of parameter space, which can be tested by physiological experiments on visual long-term memory. Such a coexistence was also found for a correlated-attractor model with block-diagonalized synaptic connections [12].

2. Model and analyses

2.1. The definitions of model

We consider an N-neuron network defined by the synaptic connections

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

$$A_{\mu\nu} = \delta_{\mu\nu} + a_1\delta_{\mu-1,\nu} + a_2\delta_{\mu+1,\nu} \tag{2}$$

with p random memory patterns $\boldsymbol{\xi}^\mu \in \{-1, 1\}^N$ ($\mu = 1, \ldots, p$). Thus the synaptic connections are decomposed into the terms which are diagonal (auto-correlation terms) or off-diagonal (cross-correlation terms) in the pattern indices. In the above equations, the memory patterns are to be cyclically identified. The time evolution equations for the output variables x_i of neurons are given by

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_i + F(h_i) \qquad i = 1, \ldots, N \tag{3}$$

where F denotes an end-cutoff nonmonotonic response function defined by [6]

$$F(x) = \begin{cases} 1 & 0 < x < \theta \\ -1 & -\theta < x < 0 \\ 0 & |x| > \theta \end{cases} \tag{4}$$

and h_i is the local field for the ith neuron,

$$h_i = \sum_{j\neq i} J_{ij}x_j. \tag{5}$$

In equation (2), the auto-correlation term, i.e. the first term, describes pattern recall and tends to stabilize each memory pattern under the dynamics defined by equation (3). On the other hand, the cross-correlation terms, i.e. the second and third terms, describe sequence processing in which the magnitudes of a_1 and a_2 determine the intensity of flows in two opposite directions, $\mu \to \mu + 1$ and $\mu \to \mu - 1$, in the state space. Griniasty *et al* [3] studied the fixed-point attractors when the synaptic connections are symmetric, i.e. $a_1 = a_2 \equiv a$, whereas Whyte *et al* [13] investigated the sequence retrieval for the asynchronous dynamics when $a_1 \neq a_2$. Both studies were made for networks of two-state neurons at finite temperatures. So their

studies at zero temperature correspond to the present study for a sigmoid response function, or for $\theta \to \infty$. Our primary interests concern the effects of nonmonotonicity on the retrieval of both correlated and Hopfield-type attractors.

2.2. Finite loading case

We introduce 2^p sublattices each of which is specified by a p-dimensional vector $\boldsymbol{X} = (\xi^1, \ldots, \xi^p)$ [14–16]. Since the number of embedded patterns is finite, each sublattice contains an infinite number of neurons when $N \to \infty$. In the thermodynamic limit, it is easy to derive the following time evolution equations for the pattern overlaps m_μ:

$$\frac{\mathrm{d}m_\mu}{\mathrm{d}t} = -m_\mu + \sum_{\boldsymbol{X}\in H^p} r(\boldsymbol{X})\xi^\mu F\left(\sum_{\nu=1}^{p} m_\nu[\xi^\nu + a_1\xi^{\nu+1} + a_2\xi^{\nu-1}]\right) \tag{6}$$

$$m_\mu = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu x_i. \tag{7}$$

Here $H^p = \{-1, 1\}^p$ and $r(\boldsymbol{X})$ denotes the probability that $\boldsymbol{\xi}_i$ coincides with p-dimensional vector $\boldsymbol{X}$ for given neuron. For random memory patterns, $r(\boldsymbol{X}) = 2^{-p}$. In the numerical simulations shown later, we primarily solve equation (6) instead of equation (3).

2.3. Extensive loading case: the generalized SCSNA

Assuming that loading rate $\alpha = p/N$ is finite, we derive the order parameter equations that describe the equilibrium states of equation (3) given by

$$x_i = F\left(\sum_{j\neq i} J_{ij}x_j\right) \qquad i = 1, \ldots, N. \tag{8}$$

Although the matrix $\boldsymbol{A}$ in equation (1) is assumed to be symmetric for simplicity, the following results are easily extended to an asymmetric case.

Let $\boldsymbol{e}^\mu$ be a set of p-dimensional normalized eigenvectors of $\boldsymbol{A}$. We introduce a set of rotated memory patterns, $\bar{\boldsymbol{\xi}}_i = (\bar{\xi}_i^1, \bar{\xi}_i^2, \ldots, \bar{\xi}_i^p)$, as

$$\boldsymbol{\xi}_i = \boldsymbol{T}\bar{\boldsymbol{\xi}}_i \tag{9}$$

$$\boldsymbol{T} = (\boldsymbol{e}^1, \boldsymbol{e}^2, \ldots, \boldsymbol{e}^p). \tag{10}$$

Each component $\bar{\xi}_i^\mu$ is statistically dependent (independent) with respect to μ (i, respectively) and satisfies the orthogonality condition,

$$\mathrm{E}[\bar{\xi}_i^\mu \bar{\xi}_i^\nu] = \delta_{\mu\nu}. \tag{11}$$

In terms of the rotated patterns, the synaptic connection in equation (1) is rewritten as

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \kappa^\mu \bar{\xi}_i^\mu \bar{\xi}_j^\mu \tag{12}$$

where κ^μ is an eigenvalue of matrix $\boldsymbol{A}$ for eigenvector $\boldsymbol{e}^\mu$. We consider the case where the equilibrium state $\boldsymbol{x}$ has nonzero overlaps $\bar{m}_\mu = \frac{1}{N}\sum_i \bar{\xi}_i^\mu x_i$ with s rotated memory patterns $\bar{\boldsymbol{\xi}}^\mu$ ($\mu = 1, \ldots, s$, $s \sim \mathrm{O}(1)$).

Following the SCSNA proposed by Shiino and Fukai [6, 17], we can easily obtain the SCSNA order parameter equation and self-consistency equation for an effective response

function Y:

$$\bar{m}_\mu = \int \mathrm{D}z \langle \bar{\xi}^\mu Y(z; \bar{\xi}^1, \bar{\xi}^2, \ldots, \bar{\xi}^s) \rangle_{\bar{\xi}} \tag{13}$$

$$q = \int \mathrm{D}z \langle (Y(z; \bar{\xi}^1, \bar{\xi}^2, \ldots, \bar{\xi}^s))^2 \rangle_{\bar{\xi}} \tag{14}$$

$$U = \frac{1}{\sigma} \int \mathrm{D}z z \langle Y(z; \bar{\xi}^1, \bar{\xi}^2, \ldots, \bar{\xi}^s) \rangle_{\bar{\xi}} \tag{15}$$

$$\mathrm{D}z \equiv \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{16}$$

$$Y(z; \bar{\xi}^1, \bar{\xi}^2, \ldots, \bar{\xi}^s) = F\left(\sum_{\mu=1}^{s} \kappa^\mu \bar{\xi}^\mu \bar{m}_\mu + \Gamma Y(z; \bar{\xi}^1, \bar{\xi}^2, \ldots, \bar{\xi}^s) + \sqrt{\alpha r} z \right) \tag{17}$$

$$r = q \int_0^1 \mathrm{d}u \frac{\kappa(u)^2}{(1-\kappa(u)U)^2} = \frac{q}{\alpha N} \mathrm{Tr}\left(\frac{\boldsymbol{A}^2}{(\boldsymbol{I} - \boldsymbol{A}U)^2} \right) \tag{18}$$

$$\Gamma = \alpha \int_0^1 \mathrm{d}u \frac{\kappa(u)^2 U}{1-\kappa(u)U} = \frac{1}{N} \mathrm{Tr}\left(\frac{\boldsymbol{A}^2 U}{\boldsymbol{I} - \boldsymbol{A}U} \right) \tag{19}$$

where $\langle \cdots \rangle_{\bar{\xi}}$ implies averaging over the condensed patterns $\bar{\boldsymbol{\xi}} = (\bar{\xi}^1, \bar{\xi}^2, \ldots, \bar{\xi}^s)$. The eigenvalue function $\kappa(u)$ in equations (18) and (19) is defined as $\kappa(\frac{\mu}{p})\kappa^\mu$ for $p, N \to \infty$. We remark that the analytical expressions of r and Γ given by equations (18) and (19) depend only on the matrix $\boldsymbol{A}$ but not explicitly on the retrieval patterns $\bar{\xi}^\mu$. This fact leads to the following order parameter equations for the equilibrium state having nonzero overlaps $m_\mu = \frac{1}{N} \sum_i \xi_i^\mu x_i$ with s original memory patterns:

$$m_\mu = \int \mathrm{D}z \langle \xi^\mu Y(z; \boldsymbol{\xi}) \rangle_\xi \tag{20}$$

$$q = \int \mathrm{D}z \langle (Y(z; \boldsymbol{\xi}))^2 \rangle_\xi \tag{21}$$

$$U = \frac{1}{\sigma} \int \mathrm{D}z z \langle Y(z; \boldsymbol{\xi}) \rangle_\xi \tag{22}$$

$$Y = F(\boldsymbol{\xi} \cdot \boldsymbol{A}_{cc} \boldsymbol{m} + \Gamma Y + \sqrt{\alpha r} z) \tag{23}$$

where $\langle \cdots \rangle_\xi$ implies averaging over the condensed patterns $\boldsymbol{\xi} = (\xi^1, \xi^2, \ldots, \xi^s)$. In equation (23), $\boldsymbol{A}_{\mathrm{CC}}$ is a submatrix of $\boldsymbol{A}$ acting on the s-dimensional space of condensed patterns. The vectors $\boldsymbol{\xi} = (\xi^1, \xi^2, \ldots, \xi^s)$ and $\boldsymbol{m} = (m_1, m_2, \ldots, m_s)$ are defined in the condensed pattern space. Note that the off-diagonal terms between the condensed and uncondensed pattern spaces can be neglected, if s is taken to be sufficiently large.

Since the matrix $A_{\mu\nu}$ in equation (2) has a translational invariance with respect to the pattern indices μ and ν, the eigenvalue function $\kappa(u)$ for $a_1 = a_2 = a$ is given as,

$$\kappa(u) = 1 + 2a\cos(\pi u) \qquad 0 \leqslant u < 1.$$

From the eigenfunctions, we obtain

$$\Gamma = \alpha \left(-\frac{1}{U} + \frac{1}{U\sqrt{1-y^2}} + \frac{1}{(1-U)\sqrt{1-y^2}} - 1 \right) \qquad y = \frac{2aU}{1-U} \tag{24}$$

$$r = q \left[\frac{1}{(1-U)^2\sqrt{1-y^2}} + \frac{4a^2}{(1-U)^4(\sqrt{1-y^2})^3} - \frac{1-\sqrt{1-y^2}}{U^2\sqrt{1-y^2}} \right]. \tag{25}$$

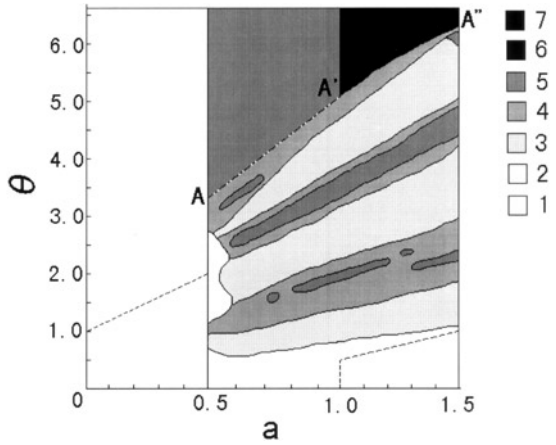

Figure 1. The number of the memory patterns which have nonvanishing positive overlaps with a correlated attractor. The synaptic connections are symmetric. For simplicity in making this diagram in the a–θ plane, a memory pattern is regarded as having a vanishing overlap if the value of the overlap was less than 2^{-7}. The attractor can be correlated only with a single memory pattern in the region above the dashed line for $a < 0.5$ and the region delimited by the dashed lines for $a > 1$ (see equations (27), (29) and (30) for details). Above the dot-dashed line AA′ given by equation (28), the attractor coincides with the correlated attractor obtained for a sigmoid response function.

3. Results

3.1. The correlated fixed-point attractors in the finite loading case

Here equation (6) is numerically solved under symmetric synaptic connections: $a_1 = a_2 = a$. In a conventional model of two-state neurons at zero temperature, i.e. in the case of $\theta \to \infty$, correlated attractors appear for $0.5 < a < 1$ [3]. For $a < 0.5$ the contributions of the auto-correlation terms to h_i completely determine its sign, thus giving only the Hopfield-type attractors. For $a > 1$, the contributions of the cross-correlation terms to h_i dominate those of the auto-correlation terms, only giving a symmetric attractor. Each correlated attractor has overlaps with a central memory pattern and four neighbouring memory patterns in both directions with increasing and decreasing pattern indices. In the present numerical simulations, the integer p was varied within $p \leqslant 25$. Since most of the results were essentially unchanged for sufficiently large values of p, the results will be shown only for $p = 14$. Note that throughout this paper we count only the central and neighbouring memory patterns in either direction of pattern indices, whenever we calculate the number of memory patterns having nonvanishing overlaps with an attractor.

To get some insight into the correlated attractors appearing in the nonmonotonic neural network, we study the number of memory patterns having nonvanishing positive overlaps with an attractor, while varying the values of θ and a. It was shown for the zero-temperature Ising-type network that the values of overlap are given as [18] $m = (1/2^7)(0, 0, 1, 3, 13, 51, 77, 51, 13, 3, 1, 0, 0, \ldots)$ for a correlated attractor with central pattern $\{\xi_i^7\}$. Therefore, we regard a memory pattern as possessing a finite overlap with an attractor if the overlap is greater than $2^{-7} = 0.0078$. Here the memory patterns having negative overlaps are not distinguished from those having vanishing overlaps.

The results are summarized in figure 1. When the value of θ is sufficiently large, the effects of nonmonotonicity disappear. In this case, the Hopfield-type attractors which are correlated

only with one of the memory patterns should be obtained for $a < 0.5$. Let $m_1 \equiv m$ be the only nonvanishing pattern overlap for a Hopfield-type attractor. Then, m must satisfy the following relation which is obtained from (6) by setting $\mathrm{d}m/\mathrm{d}t = 0$ and $m_2 = \cdots = m_p = 0$:

$$m = \tfrac{1}{4}\{2F(m) + F((1-2a)m) + F((1+2a)m)\}. \tag{26}$$

The first, second and third terms in parenthesis appear from the sublattices with $\xi^2 = -\xi^p$, $\xi^1 = -\xi^2 = -\xi^p$ and $\xi^1 = \xi^2 = \xi^p$, respectively. The equation (26) has a unique solution $m = 1$ when the maximum value of the local fields is less than the value of θ. Thus

$$m = 1 \qquad \text{for} \quad \theta > 1 + 2a \tag{27}$$

in the region above the dashed line drawn in figure 1 for $a < 0.5$. However, when $\theta < 1 + 2a$, $m = 1$ cannot be a solution of (26) any more, since the last term in the right-hand side vanishes for the values of m. The behaviour of the attractor in this case will be studied later.

In the case where $a > 0.5$, the correlated attractors appear and accordingly the figure shows complicated behaviours. When $0.5 < a < 1$, the attractor is correlated with five (i.e. a central + four neighbouring) memory patterns for (a, θ) located above some line connecting the points A and A′ (see equation (28) for details). When $a > 1$, the attractor of this neural network approaches a symmetric state correlated with all memory patterns, as θ becomes large in the region above some curve connecting A′ and A″. These results are consistent with those for a sigmoid response function [3, 19]. In contrast, in the region below the curve AA′A″, the model shows complicated behaviour. In this region, the attractor always has nonvanishing overlaps with more than one memory pattern for $0.5 < a < 1$. On the other hand, for $a > 1$ the attractor has a nonvanishing overlap only with a single memory pattern in the region delimited by the dashed lines (see equations (29) and (30) for details). Except for this difference, the figure suggests that the correlated attractor shows similar behaviour for both $0.5 < a < 1$ and $a > 1$. However, as we will see below, this is not indeed the case.

To see the differences in the behaviour of the correlated attractor in the three cases that $a < 0.5$, $0.5 < a < 1$ and $a > 1$, the distributions of pattern overlaps over memory patterns are numerically calculated by varying the value of θ. The results are shown in figures 2(*a*)–(*c*) for $a = 0.25$, $a = 0.9$ and $a = 1.1$, respectively. In all of the figures, the attractor which is maximally correlated with $\{\xi_i^7\}$ is shown.

Figure 2(*a*) proves that for $a < 0.5$ the attractor is correlated only with a single memory pattern as far as $\theta > 1 + 2a = 1.5$, as was shown in (27). As the value of θ is decreased below this critical line, the attractor starts to have negative or positive overlaps with other memory patterns.

Figures 2(*b*) and (*c*) show clear differences in the behaviour of correlated attractors for $0.5 < a < 1$ and $a > 1$. In the case where $a = 0.9$, the shape of the distribution changes moderately as θ is decreased or the nonmonotonicity is increased. The correlated attractor at any value of θ has nonvanishing overlaps with a relatively small number of memory patterns, which is usually less than or equal to five. The critical line AA′, above which the correlated attractors coincide with those obtained for a sigmoid response function, is easily obtained. As shown previously, the values of pattern overlaps for the correlated attractors are given as $m = (1/2^7)(0, 0, 1, 3, 13, 51, 77, 51, 13, 3, 1, 0, 0, \ldots)$ [18]. Then the local fields given by equation (5) take the maximum value $(1+2a)(1+3+13+51+77+51+13+3+1)/2^7 = (1+2a)\frac{213}{128}$ in the sublattices with $\xi^3 = \xi^4 = \cdots = \xi^{11} = 1$. To have the same correlated attractors as obtained for a sigmoid response function, this maximum value should be less than the value of θ. Thus the critical line AA′ is given by

$$\theta = \tfrac{213}{64}a + \tfrac{213}{128} \qquad 0.5 < a < 1. \tag{28}$$

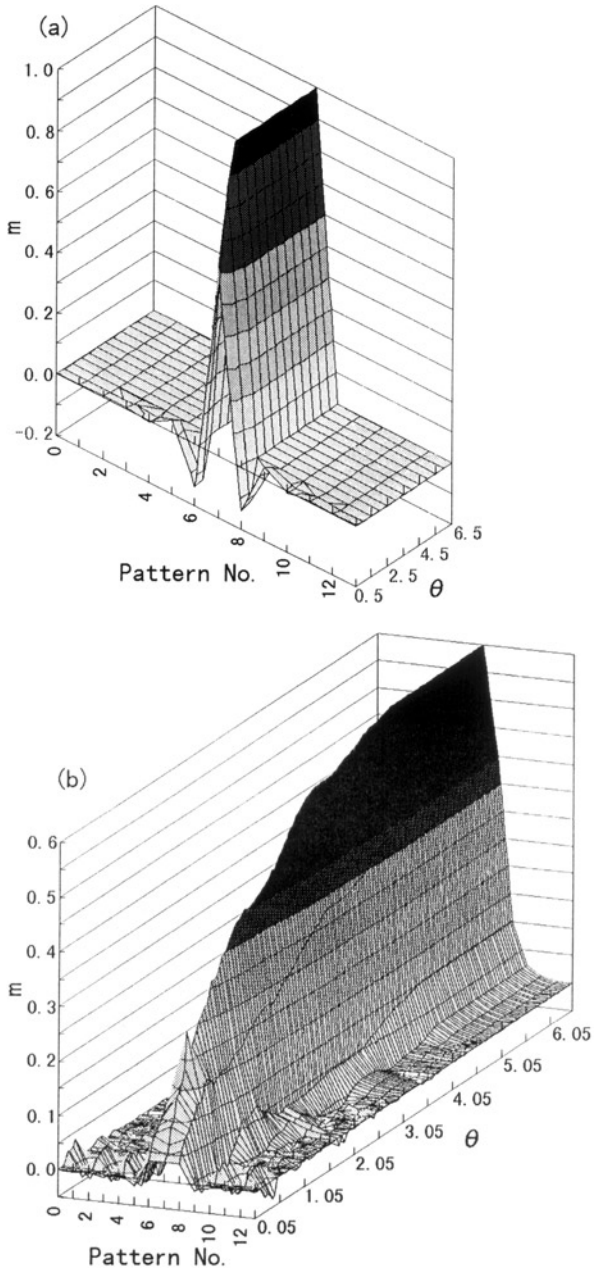

Figure 2. The overlap distributions over memory patterns for various values of θ, when (*a*) $a = a_1 = a_2 = 0.25$, (*b*) 0.9 and (*c*) 1.1. The total number of embedded memory patterns is given by $p = 14$.

In contrast to the above case, when the shape of the distribution changes dramatically at some values of θ, as its value is decreased from a large value. The first dramatic change in the shape occurs in the neighbourhood of $\theta \approx 5$, when the value of θ crosses the curve A′A″ in the diagram shown in figure 1. Near this value of θ, the width of distribution, or the number

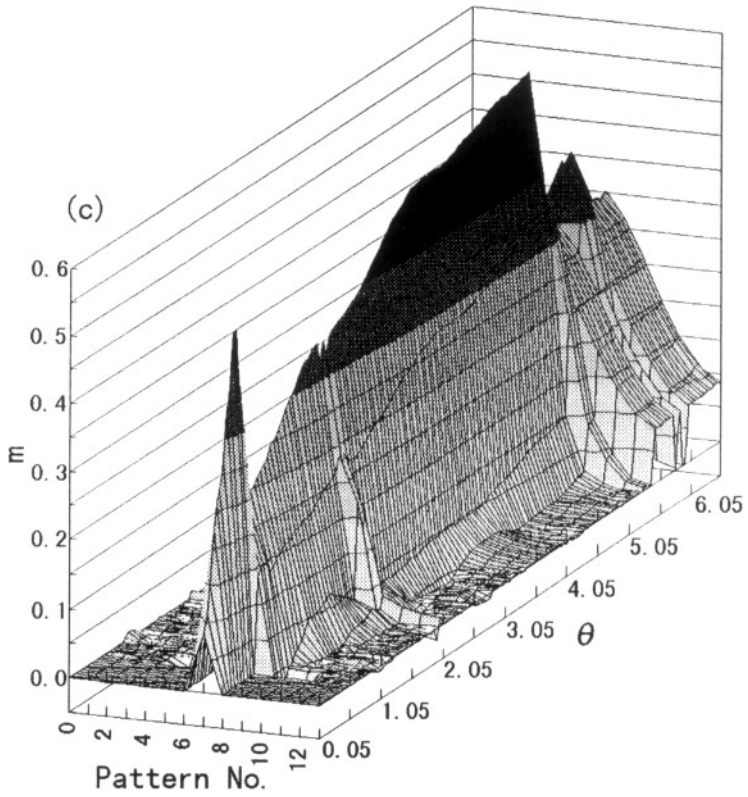

Figure 2. (Continued)

of memory patterns correlated with the attractor, is decreased rapidly. Below this value, the shape of distribution does not show significant changes until $\theta \approx 0.6$, except the changes that occur in the neighbourhood of $\theta \approx 2$. Near $\theta \approx 0.6$, the attractor begins to be correlated only with the central memory pattern, which results in a rapid increase in the peak height of the distribution. This behaviour can be understood from equation (26) as follows. Note that $1 - 2a < 0$ for $a > 1$. If $1 - 2a < -\theta$, $m = \frac{1}{2}$ can be a solution to equation (26) with $F(m) = 1$, and it remains to be a solution as far as $\theta > \frac{1}{2}$. Thus,

$$m = \tfrac{1}{2} \qquad \text{for} \quad \tfrac{1}{2} < \theta < a - \tfrac{1}{2}. \tag{29}$$

For $\theta < \frac{1}{2}$, the equation $2m = F(m)$ must be satisfied, and a solution is given by

$$m = \theta. \tag{30}$$

For very small values of θ, the numerically-obtained attractors exhibited the oscillatory instabilities around the values given above.

The above results for $a = 0.9$ and $a = 1.1$ reveal that a retrieval state of the neural network exhibits quite different behaviours for $0.5 < a < 1$ and $a > 1$. They also reveal that the complicated features shown in the region of figure 1 below the curve AA′A″ are just the artifact arising from the criteria employed for finding nonvanishing overlaps. It is noted that for a highly nonmonotonic response function with a small value of θ, the retrieval state can be correlated only with a single memory pattern when $a > 1$, whereas it is always correlated with more than one memory patterns when $0.5 < a < 1$.

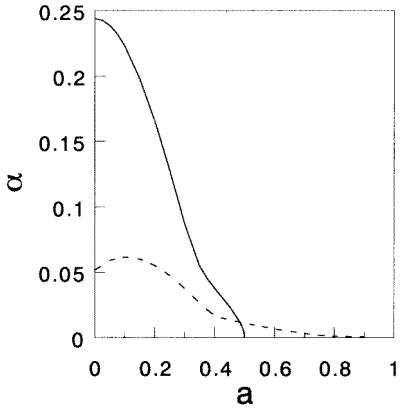

Figure 3. The phase diagram obtained analytically by SCSNA for $s = 13$. The solid curve represents the capacities for the Hopfield-type attractors, while the dashed curve those for the correlated attractors.

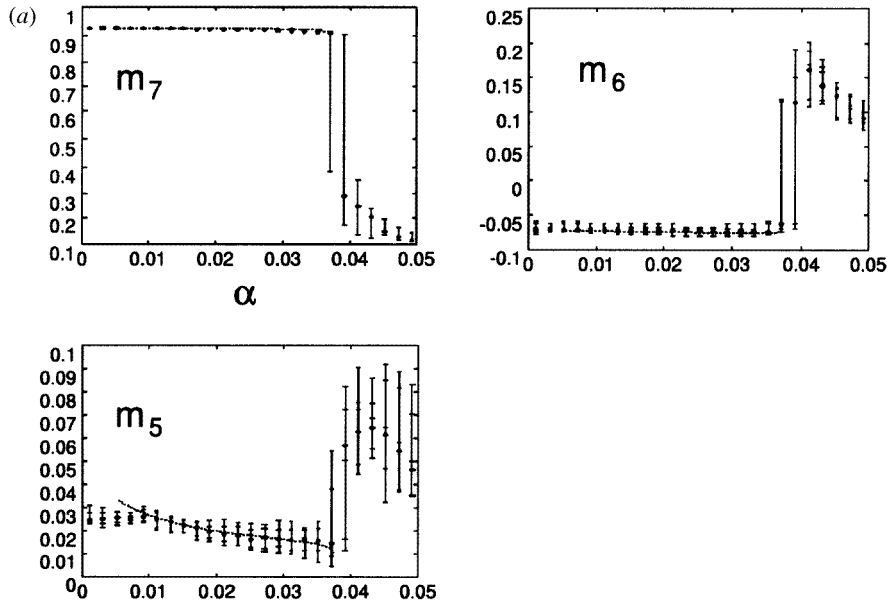

Figure 4. The results of numerical simulations are shown in terms of pattern overlaps for (*a*) Hopfield-type attractors and (*b*) correlated attractors at $a = 0.4$. The central pattern which has the largest overlap with the attractors is given by $\{\xi_i^7\}$. The dashed curves represent the values obtained theoretically by SCSNA.

3.2. Phase diagram in the extensive loading case

To obtain the phase diagram, we solve equations (20)–(25) numerically, while varying the values of a and α. In the finite loading case, the nonmonotonicity of response function resulted in the occurrence of negatively-correlated attractors below the line $\theta = 1 + 2a$ for $a < 0.5$. It will be interesting to see whether the retrieval properties qualitatively change when the parameter values across this line in the (a, θ) space in the extensive loading case. Therefore the value of θ is fixed as $\theta = 1.5$. The resultant phase diagram is shown in figure 3 for $s = 13$. Several results shown in this phase diagram are expected from those obtained for a sigmoid response function: (i) the correlated attractors exist only for $a < 1$; (ii) the uncorrelated Hopfield-type attractors exist only for $a < 0.5$; (iii) outside the above two types

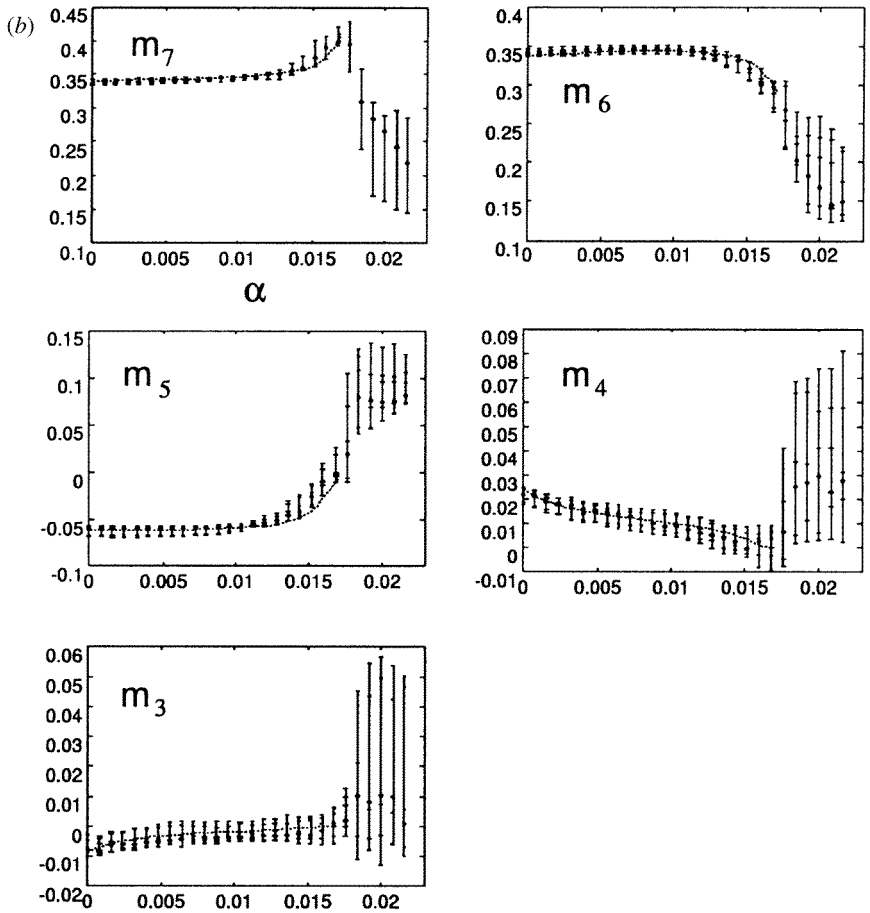

Figure 4. (Continued)

of retrieval phases are the spin glass states. Besides the spin glass states, a symmetric state in which all memory patterns are symmetrically mixed appears for finite s. However, the state is not presented in the phase diagram since it should not appear at the limit $s \to \infty$; (iv) the memory capacities for the correlated attractors are much smaller than those for the Hopfield-type attractors. Also the capacities for the Hopfield-type attractors are enhanced by the nonmonotonicity of response functions [6].

However, the phase diagram in figure 3 exhibits a qualitatively different feature which was not shown in [19]. Namely, the Hopfield-type and correlated attractors coexist for $a < 0.5$. This coexistence of different attractors has been found also for a sigmoid or step-function-like response function (results not shown). In numerically solving the order parameter equations for $a < 0.5$, the convergence to the correlated-attractor solutions is very sensitive to the choices of initial values for iterations. Therefore these solutions could easily be missed in the numerical analysis.

To confirm the above results of analytic studies, we conducted numerical simulations for several cases. Since the critical loading rate can be rather small for the correlated attractors, N should be, in general, taken to be very large numbers. We employed a maximum of 350 000 neurons in the simulations. For this value of N, $s = 13$ corresponds only to 3.7% of the

critical pattern number $\alpha_c N$ at $\alpha_c = 0.001$. Figure 4(*a*) shows the values of pattern overlaps for the Hopfield-type attractors at $a = 0.4$, when the target pattern is $\{\xi^7\}$. It is clear from the figure that the numerically obtained values coincide well with those obtained theoretically (dashed curves) for $\alpha < \alpha_c$. At $\alpha = \alpha_c$, the numerically obtained values rapidly jump, which signals the occurrence of phase transition. In figure 4(*b*), the results for the correlated attractors are shown at $a = 0.4$. The presence of the attractors below $\alpha \approx 0.017$ is clearly seen from the figures, which further proves the good coincidence between the theoretical and numerical results. Thus, the results of numerical simulations confirm the coexistence of correlated and Hopfield-type attractors for $a < 0.5$.

4. Conclusion

In the present paper, we have investigated the properties of the correlated attractors in an associative memory model with a nonmonotonic response function in the cases of both finite and extensive memory loading. For a finite number of memory patterns, we have shown that the number of memory patterns correlated with an attractor can change dramatically with the degree of the nonmonotonicity. For an infinite number of memory patterns, we have extended the applicability of the self-consistent signal-to-noise analysis and derived the analytic expressions for order parameters. The phase diagram obtained from the analysis revealed the coexistence of the Hopfield-type attractors and correlated ones for $a < 0.5$, when the loading rate α is small. Since this coexistence was also found for a sigmoid response function, and the correlated-attractor neural network was originally proposed as a model of long-term visual memory in the monkey infero-temporal area, it seems interesting to examine whether the two different types of attractors indeed coexist in the brain area.

Acknowledgments

This work was partially supported by Grant-in-Aid for Scientific Research on Priority Areas No 09268235 from the Ministry of Education, Japan.

References

[1] Miyashita Y and Chang HS 1988 Neuronal correlate of pictorial short-term memory in the primate temporal cortex *Nature* **331** 68–70
[2] Miyashita Y 1988 Neuronal correlate of visual associative long-term memory in the primate temporal cortex *Nature* **335** 817–20
[3] Griniasty M, Tsodyks M V and Amit D J 1993 Conversion of temporal correlations between stimuli to spatial correlations between attractors *Neural Comp.* **5** 1–17
[4] Amit D J, Brunel N and Tsodyks MV 1994 Correlations of cortical Hebbian reverberations: theory versus experiment *J. Neurosci.* **14** 6435–45
[5] Morita M 1993 Associative memory with nonmonotone dynamics. *Neural Net.* **6** 115–26
[6] Shiino M and Fukai T 1993 Self-consistent signal-to-noise analysis of the statistical behaviour of analog neural networks and enhancement of the storage capacity *Phys. Rev.* E **48** 867–97
[7] Yoshizawa S, Morita M and Amari S 1993 Capacity of associative memory using a nonmonotonic neuron model *Neural Net.* **6** 167–76
[8] Boffetta G, Monasson R and Zecchina R 1993 Symmetry breaking in non-monotonic neural networks *J. Phys. A: Math. Gen.* **26** L507–13
[9] Brunel N and Zecchina R 1994 Response functions improving performance in analog attractor neural networks *Phys. Rev.* E **49** R1823–6
[10] Meilijson I and Ruppin E 1993 History-dependent attractor neural networks *Network* **4** 195–221
[11] Meilijson I and Ruppin E 1994 Optimal signalling in attractor neural networks *Network* **5** 277–98

[12] Shiino M 1997 Associative memory neural networks with asymmetric connections *Int. Symp. on Nonlinear Theory and its Applications (NOLTA'97, Honolulu, USA)* p 665
[13] Whyte W, Sherrington D and Coolen ACC 1995 Competition between pattern recall and sequence processing in a neural network storing correlated patterns *J. Phys. A: Math. Gen.* **28** 3421–37
[14] von Hemmen J L 1982 Classical spin-glass model *Phys. Rev. Lett.* **49** 409–12
[15] Riedel U, Kuhn R and von Hemmen J L 1988 Sequences and chaos in neural nets *Phys. Rev.* A **38** 1105–8
[16] Fukai T and Shiino M 1990 Asymmetric neural networks incorporating the Dale hypothesis and noise-driven chaos *Phys. Rev. Lett.* **64** 1465–8
[17] Mimura K, Okada M and Kurata K 1998 Associative memory model with forgetting process using nonmonotonic neurons *IEICE Trans. Inf. Syst.* **E81-D** 1298–304
[18] Cugliandolo L F 1994 Correlated attractors from uncorrelated stimuli *Nuovo Cimento* **6** 220–4
[19] Cugliandolo L F and Tsodyks M V 1994 Capacity of networks with correlated attractors *J. Phys. A: Math. Gen.* **27** 741–56