Bistability of mixed states in a neural network storing hierarchical patterns

2000 J. Phys. A: Math. Gen. 33 2725

(http://iopscience.iop.org/0305-4470/33/14/308)

J. Phys. A: Math. Gen. **33** (2000) 2725–2737. Printed in the UK PII: S0305-4470(00)08219-6

Bistability of mixed states in a neural network storing hierarchical patterns

Kaname Toya†, Kunihiko Fukushima‡,Yoshiyuki Kabashima§ and Masato Okada‖¶

† Department of Systems and Human Science, Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama, Toyonaka, Osaka 560-8531, Japan
‡ Department of Information and Communication Engineering, The University of Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan
§ Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama 226-8502, Japan
‖ Kawato Dynamic Brain Project, Japan Science and Technology Corporation, 2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

E-mail: masato@erato.atr.co.jp

Received 28 September 1999, in final form 19 January 2000

Abstract. We discuss the properties of equilibrium states in an autoassociative memory model storing hierarchically correlated patterns (hereafter, hierarchical patterns). We will show that symmetric mixed states (hereafter, mixed states) are bistable on the associative memory model storing the hierarchical patterns in a region of the ferromagnetic phase. This means that the first-order transition occurs in this ferromagnetic phase. We treat these contents with a statistical mechanical method (SCSNA) and by computer simulation. Finally, we discuss a physiological implication of this model. Sugase *et al* (1999 *Nature* **400** 869) analysed the time-course of the information carried by the firing of face-responsive neurons in the inferior temporal cortex. We also discuss the relation between the theoretical results and the physiological experiments of Sugase *et al*.

1. Introduction

There are two kinds of hypotheses regarding internal representations of memory items in the brain. First, there is the 'distributed representation' hypothesis, which assumes that our memory items are encoded by neuron activity patterns. But another type of hypothesis (e.g., the 'grandmother cell' hypothesis) has been proposed, where the memory items are represented by the excitation of corresponding neurons: that is, a 'local representation' hypothesis. Let us consider a neural network model consisting of N neurons. The number of memory items in the local representation is $\mathrm{O}(N)$. One might think that the number of memory items in the distributed representation is $\mathrm{O}(2^N)$ in the case of storing binary patterns. However, this consideration is meaningless. If we consider error correcting abilities, this becomes $\mathrm{O}(N)$ on the basis of the statistical mechanical theories (for example [2]). One of the remarkable advantages of the distributed representation is that the relationship between the memory items is naturally implemented by the distance of memory patterns. However, many studies on associative memory models have been confined to those with uniformly distributed patterns.

¶ Author to whom correspondence should be addressed.

0305-4470/00/142725+13$30.00

Thus, we should discuss a model that stores some structural patterns. We will extend the conventional SCSNA [3] to a generalized SCSNA in order to treat a model with any structural pattern and a general class of output functions of neuron.

When an associative memory model is made to store memory patterns as a result of correlation learning, a pattern, which has an uniform overlap with some of the stored memory patterns, automatically becomes the equilibrium state of the model. This is called the mixed state. It is not appropriate to think that this mixed state is a side-effect and/or that it is unnecessary for information processing. Amari has discussed a 'concept formation' using the stability of mixed state [4]. The correlated attractor [5, 6] accounting for the physiological experiments by Miyashita [7] can be interpreted as a mixed state in a broad sense. Recently, Parga and Rolls [8] used the mixed state in their research on the mechanism of invariant recognition with a coordinate transformation in the visual system.

The aim of this research is to study the properties of mixed states in the autoassociative memory model storing hierarchical patterns by utilizing the generalized SCSNA and computer simulation. First, we discuss a model that stores patterns in which a two-stage hierarchy exists. With the generalized SCSNA, we derive the order parameter equations for the equilibrium state in this model. By solving the obtained equations, we show that two kinds of mixed states coexist in a particular region of the retrieval phase (ferromagnetic phase). It is a characteristic of the two kinds of mixed states that they have different values of cross-talk noise variances, that is, they are influenced by the uncondensed patterns in different ways. This kind of bistability in the retrieval phase has not been previously reported. We will show that the bistability of mixed states does not depend on the number of patterns in the same cluster. Generally speaking, it is not so easy to confirm the multi-stable states by computer simulation. However, we succeed in confirming the bistability of the mixed states by computer simulation after considering the qualitative properties of the retrieval dynamics in this model. Next, we treat a model storing a set of hierarchical patterns and uniformly distributed patterns in order to investigate universality of the bistability. From these theoretical results, we found that a bistability of mixed states also exists in this model. Contrary to these models, such bistability of mixed states does not exist in a model where the contribution of uncondensed patterns to the synaptic couplings is replaced by the spin glass type interaction. We will discuss the reason for this by using the SCSNA.

Recently, Sugase *et al* have reported interesting phenomena concerning the temporal dynamics of face-responsive neurons in the inferior temporal (IT) cortex [1]. In the discussion, we will examine the relation between the obtained results and the physiological experiments by Sugase *et al*.

2. Model

Let us consider a recurrent neural network consisting of N neurons with an output function $F(\cdot)$. We employ the synchronous dynamics,

$$x_i^{t+1} = F\left(\sum_{j\neq i}^{N} J_{ij}x_j^t\right) \tag{1}$$

where x_i^t represents a state of the ith neuron at discrete time t, and discuss the case of the thermodynamics limit ($N \to \infty$). J_{ij} in the above equation denotes a synaptic coupling from the jth neuron to the ith neuron. In this work, we discuss the equilibrium state of equation (1). For simplicity, we treat a two-stage hierarchy, which is one of the simplest cases. This can be easily extended to more complex hierarchies. One can use many procedures to make the set

of ultrametric patterns, but we employ the following method. Each component $\xi_i^{\mu,\nu}$ of pattern $\boldsymbol{\xi}^{\mu,\nu}$ is a random variable drawn from the following probability distributions:

$$P[\xi_i^{\mu} = \pm 1] = \tfrac{1}{2} \qquad i = 1, 2, \ldots, N \qquad \mu = 1, 2, \ldots, p \tag{2}$$

$$P[\xi_i^{\mu,\nu} = \pm 1] = \frac{1 \pm b\xi_i^{\mu}}{2} \qquad \nu = 1, 2, \ldots, s \tag{3}$$

with $0 \leqslant b \leqslant 1$. The distance between patterns $\boldsymbol{\xi}^{\mu,\nu}$ is expressed by

$$E[\xi_i^{\mu,\nu}\xi_i^{\mu',\nu'}] = \delta_{\mu\mu'}(\boldsymbol{B})_{\nu\nu'} \tag{4}$$

$$(\boldsymbol{B})_{\nu\nu'} \equiv \delta_{\mu\mu'}(\delta_{\nu\nu'} + b^2(1 - \delta_{\nu\nu'})) \tag{5}$$

where $E[\cdot]$ stands for an average with respect to the probabilities distributed in equations (2) and (3), and $\delta_{\mu\mu'}$ is the Kronecker's δ defined as

$$\delta_{\mu\mu'} = \begin{cases} 1 & (\mu = \mu') \\ 0 & (\mu \neq \mu'). \end{cases} \tag{6}$$

According to the definition of the matrix $\boldsymbol{B}$ in equation (5), $\boldsymbol{B}$ is $s \times s$ matrix. As shown in equation (4), the memory patterns $\boldsymbol{\xi}^{\mu,\nu}$ have a two-stage ultrametric structure. $(\boldsymbol{B})_{\nu\nu'}$ in equation (4) stands for the element in the νth row, the ν'th column of matrix $\boldsymbol{B}$. $\boldsymbol{\xi}^{\mu,\nu}$ are ps uniformly generated patterns when $b = 0$, while $\boldsymbol{\xi}^{\mu,\nu}$ in the μth cluster are the same when $b = 1$.

We employ the simple Hebbian rule as the learning rule, and the synaptic coupling J_{ij} is set to

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{\alpha N}\sum_{\nu=1}^{s}\xi_i^{\mu,\nu}\xi_j^{\mu,\nu} \tag{7}$$

where $\alpha = \frac{p}{N}$. Since the number of clusters is αN, we call α the loading rate. $\boldsymbol{\xi}^{\mu}$ is not explicitly used for the learning rule. We discuss two kinds of models. One is the previously defined model with the couplings J_{ij} given by equation (7), which we call model 1. The other model, model 2, is shown here with the following couplings J_{ij}:

$$J_{ij} = \frac{1}{N}\sum_{\nu=1}^{s}\xi_i^{1,\nu}\xi_j^{1,\nu} + \frac{1}{N}\sum_{\mu=2}^{\alpha N}\xi_i^{\mu}\xi_j^{\mu}. \tag{8}$$

We will examine the stability of the memory pattern and a mixed state which has a uniform overlap with an odd number of memory patterns in the same cluster. There exist mixed states which have finite overlaps with memory patterns in different clusters. However, we do *not* discuss this topic in this paper.

3. A generalized SCSNA

There are two kinds of statistical–mechanical theories treating the equilibrium properties of associative memory models. One is based on the SCSNA proposed by Shiino and Fukai [3]. The other one is the replica theory, which has been used to analyse the equilibrium states in model 1 [9]. However, the replica theory cannot be applied to a system where the free energy cannot be defined (e.g., a model with a nonmonotonic output function [10] or an oscillator associative memory model [11]), while the SCSNA can treat these systems without the energy function. Since previous studies with the SCSNA have mainly focused on systems with uniformly distributed patterns, the SCSNA cannot be directly applied to model 1.

We will extend the previously proposed SCSNA to a generalized SCSNA in this section in order to treat a model storing a set of structural patterns $\boldsymbol{\sigma}^{\rho}(\rho = 1, 2, \ldots, \hat{\alpha}N)$ randomly generated by an independent identical probability distribution with respect to i. The correlation matrix $(\boldsymbol{C})_{\rho\rho'}$ between the two patterns $\boldsymbol{\sigma}^{\rho}$ and $\boldsymbol{\sigma}^{\rho'}$ is given by

$$(\boldsymbol{C})_{\rho\rho'} \equiv \frac{1}{N}\sum_{i=1}^{\hat{\alpha}N} \sigma_i^{\rho}\sigma_i^{\rho'} \tag{9}$$

$$= E[\sigma_i^{\rho}\sigma_i^{\rho'}]. \tag{10}$$

We consider a recurrent network consisting of N neurons with the output function $F(\cdot)$ and synaptic couplings J_{ij} given by

$$J_{ij} = \frac{1}{N}\sum_{\rho=1}^{\hat{\alpha}N} \sigma_i^{\rho}\sigma_j^{\rho}. \tag{11}$$

Since the number of memory patterns is $\hat{\alpha}N$, $\hat{\alpha}$ is defined as the loading rate in this model. $\boldsymbol{\xi}^{\mu,\nu}$ or $\boldsymbol{B}$ in equation (4) is an example of $\boldsymbol{\sigma}^{\rho}$ or $\boldsymbol{C}$ in equation (10), respectively,

$$\boldsymbol{\sigma}^{\rho} \leftrightarrow \boldsymbol{\xi}^{\mu,\nu}\rho = s(\mu - 1) + \nu \tag{12}$$

$$(\boldsymbol{C})_{\rho\rho'} \leftrightarrow \delta_{\mu\mu'}(\boldsymbol{B})_{\nu\nu'} \tag{13}$$

$$\hat{\alpha} \leftrightarrow \alpha s. \tag{14}$$

We consider the case where the state of neurons at discrete time step t, $\boldsymbol{x}^t$, is synchronously updated,

$$x_i^{t+1} = F\left(\sum_{j\neq i}^{N} J_{ij}x_j^t\right) \tag{15}$$

and discuss the equilibrium state $\boldsymbol{x}$ with the limit $t \to \infty$. We introduce a set of rotated memory patterns, $\bar{\boldsymbol{\sigma}} = \{\bar{\sigma_i}^1, \bar{\sigma_i}^2, \ldots, \bar{\sigma_i}^{\hat{\alpha}N}\}$, as

$$\bar{\sigma}_i^{\rho} = \frac{1}{\sqrt{\kappa_{\rho}}}\sum_{\rho'=1}^{\hat{\alpha}N} W_{\rho\rho'}\sigma_i^{\rho'} \tag{16}$$

$$\boldsymbol{W} = (\boldsymbol{w}^1, \boldsymbol{w}^2, \ldots, \boldsymbol{w}^{\hat{\alpha}N})^T. \tag{17}$$

$\boldsymbol{w}^{\rho}(\rho = 1, 2, \ldots, \hat{\alpha}N)$ and $\kappa_{\rho}(\rho = 1, 2, \ldots, \hat{\alpha}N)$ in the above equations represent the ρth $\hat{\alpha}N$-dimensional normalized eigenvector of matrix $\boldsymbol{C}$ and the ρth eigenvalue of matrix $\boldsymbol{C}$ for $\boldsymbol{w}^{\rho}$, respectively. Each component $\bar{\sigma}_i^{\rho}$ is statistically dependent (independent) on ρ (i), respectively, and satisfies the following conditions:

$$\frac{1}{N}\sum_{i=1}^{N}(\bar{\sigma}_i^{\rho})^2 = 1 \tag{18}$$

$$\frac{1}{N}\sum_{i=1}^{N}\bar{\sigma}_i^{\rho}\bar{\sigma}_i^{\rho'} \sim \mathrm{O}\left(\frac{1}{\sqrt{N}}\right) \qquad \rho \neq \rho'. \tag{19}$$

Using the rotated patterns, we can rewrite J_{ij} in equation (11) as

$$J_{ij} = \frac{1}{N}\sum_{\rho=1}^{\hat{\alpha}N} \kappa_{\rho}\bar{\sigma}_i^{\rho}\bar{\sigma}_j^{\rho}. \tag{20}$$

The overlaps $\bar{m}^{\rho}$ between the equilibrium state $\boldsymbol{x}$ and $\bar{\boldsymbol{\sigma}}^{\rho}$ are defined by the following equation:

$$\bar{m}^{\rho} = \frac{1}{N}\sum_{i=1}^{N}\bar{\sigma}_i^{\rho}x_i. \tag{21}$$

If one assumes that the equilibrium state $\boldsymbol{x}$ has nonzero overlaps with $\bar{s}$ rotated patterns $\bar{\boldsymbol{\sigma}}^{\rho}(1 \leqslant \rho \leqslant \bar{s})$, we can derive the SCSNA order parameter equations [12] (see appendix A):

$$\bar{m}^{\rho} = \int \mathrm{D}z \langle \bar{\sigma}^{\rho} Y(z, \bar{\sigma}^{1}, \bar{\sigma}^{2}, \ldots, \bar{\sigma}^{\bar{s}}) \rangle_{\bar{\sigma}} \tag{22}$$

$$q = \int \mathrm{D}z \langle (Y(z, \bar{\sigma}^{1}, \bar{\sigma}^{2}, \ldots, \bar{\sigma}^{\bar{s}}))^{2} \rangle_{\bar{\sigma}} \tag{23}$$

$$U = \frac{1}{\sqrt{\hat{\alpha} r}} \int \mathrm{D}z \, z \langle Y(z, \bar{\sigma}^{1}, \bar{\sigma}^{2}, \ldots, \bar{\sigma}^{\bar{s}}) \rangle_{\bar{\sigma}} \tag{24}$$

$$\mathrm{D}z = \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(\frac{-z^{2}}{2}\right) \tag{25}$$

$$Y(z, \bar{\sigma}^{1}, \bar{\sigma}^{2}, \ldots, \bar{\sigma}^{\bar{s}}) = F\left(\sum_{\rho=1}^{\bar{s}} \kappa_{\rho} \bar{\sigma}^{\rho} \bar{m}^{\rho} + \Gamma Y(z, \bar{\sigma}^{1}, \bar{\sigma}^{2}, \ldots, \bar{\sigma}^{\bar{s}}) + \sqrt{\hat{\alpha} r} z \right) \tag{26}$$

$$r = q \int_{0}^{1} \mathrm{d}u \frac{\kappa(u)^{2}}{(1 - \kappa(u)U)^{2}} = \frac{q}{\hat{\alpha} N} \mathrm{Tr}\left(\frac{\boldsymbol{C}^{2}}{(\boldsymbol{I} - \boldsymbol{C}U)^{2}} \right) \tag{27}$$

$$\Gamma = \hat{\alpha} \int_{0}^{1} \mathrm{d}u \frac{\kappa(u)^{2} U}{1 - \kappa(u)U} = \frac{1}{N} \mathrm{Tr}\left(\frac{\boldsymbol{C}^{2} U}{\boldsymbol{I} - \boldsymbol{C}U} \right) \tag{28}$$

where $\langle \cdots \rangle_{\bar{\sigma}}$ stands for an average over the condensed patterns $\bar{\boldsymbol{\sigma}} = (\bar{\sigma}^{1}, \bar{\sigma}^{2}, \ldots, \bar{\sigma}^{\bar{s}})$. We can express the sum of κ_{ρ} in terms of an integration along continuous eigenvalue $\kappa(\frac{\rho}{\hat{\alpha} N}) \equiv \kappa_{\rho}$ for $p, N \to \infty$. Note that the analytical expressions of r and Γ given by equations (27) and (28) depend only on the matrix $\boldsymbol{C}$ and do not explicitly depend on the condensed patterns $\bar{\boldsymbol{\sigma}}^{\rho}(\rho = 1, 2, \ldots, \bar{s})$. This fact leads to the following order parameter equations for the equilibrium state having nonzero overlaps $m^{\rho} = \frac{1}{N} \sum_{i=1}^{N} \sigma_{i}^{\rho} x_{i}$ with s original memory patterns $\boldsymbol{\sigma}^{\rho}(\rho = 1, 2, \ldots, s)$:

$$m^{\rho} = \int \mathrm{D}z \langle \sigma^{\rho} Y(z, \sigma^{1}, \sigma^{2}, \ldots, \sigma^{s}) \rangle_{\sigma} \tag{29}$$

$$q = \int \mathrm{D}z \langle (Y(z, \sigma^{1}, \sigma^{2}, \ldots, \sigma^{s}))^{2} \rangle_{\sigma} \tag{30}$$

$$U = \frac{1}{\sqrt{\hat{\alpha} r}} \int \mathrm{D}z \, z \langle Y(z, \sigma^{1}, \sigma^{2}, \ldots, \sigma^{s}) \rangle_{\sigma} \tag{31}$$

$$\mathrm{D}z = \frac{\mathrm{d}z}{\sqrt{2\pi}} \exp\left(\frac{-z^{2}}{2}\right) \tag{32}$$

$$Y(z, \sigma^{1}, \sigma^{2}, \ldots, \sigma^{s}) = F\left(\sum_{\rho=1}^{s} \sigma^{\rho} m^{\rho} + \Gamma Y(z, \sigma^{1}, \sigma^{2}, \ldots, \sigma^{s}) + \sqrt{\hat{\alpha} r} z \right) \tag{33}$$

$$r = \frac{q}{\hat{\alpha} N} \mathrm{Tr}\left(\frac{\boldsymbol{C}^{2}}{(\boldsymbol{I} - \boldsymbol{C}U)^{2}} \right) \tag{34}$$

$$\Gamma = \frac{1}{N} \mathrm{Tr}\left(\frac{\boldsymbol{C}^{2} U}{\boldsymbol{I} - \boldsymbol{C}U} \right). \tag{35}$$

Note that the off-diagonal terms of the matrix $\boldsymbol{C}$ between the condensed and uncondensed pattern spaces can be neglected if $\bar{s}$ is taken to be sufficiently large but O(1).

4. Results

We discuss the case where $F(\cdot) = \mathrm{sgn}(\cdot)$ defined by

$$\mathrm{sgn}(u) = \begin{cases} 1 & (u \geqslant 0) \\ -1 & (u < 0). \end{cases} \tag{36}$$

First, we will show the results of the analysis for the equilibrium properties in model 1. We apply the generalized SCSNA proposed in section 3 to model 1. We rewrite equation (13) as

$$\boldsymbol{C} = \underbrace{\boldsymbol{B} \bigoplus \boldsymbol{B} \bigoplus, \ldots, \bigoplus \boldsymbol{B}}_{p}. \tag{37}$$

The matrix $\boldsymbol{B}$ has the eigenvalues $\lambda_1 = 1 + (s-1)b^2$, $\lambda_\nu = 1 - b^2 (2 \leqslant \nu \leqslant s)$. One assumes that the overlaps $m^{1,\nu} = \frac{1}{N}\sum_{i=1}^{N} \xi_i^{1,\nu} x_i (\nu = 1, 2, \ldots, s)$ have values of O(1). By considering the relations given by equations (12), (14) and (37), we obtained the following SCSNA order parameter equations for this model:

$$m^{1,\nu} = \left\langle \xi^{1,\nu} \mathrm{erf}\left(\frac{\sum_{\sigma=1}^{s} \xi^{1,\sigma} m^{1,\sigma}}{\sqrt{2\alpha r}} \right) \right\rangle_{\xi^1} \tag{38}$$

$$q = 1 \tag{39}$$

$$r = q \sum_{\nu=1}^{s} \left(\frac{{\lambda_\nu}^2}{(1 - \lambda_\nu U)^2} \right) \tag{40}$$

$$U = \sqrt{\frac{2}{\pi \alpha r}} \left\langle \exp\left(-\left(\sum_{\sigma=1}^{s} \frac{\xi^{1,\sigma} m^{1,\sigma}}{\sqrt{2\alpha r}} \right)^2 \right) \right\rangle_{\xi^1} \tag{41}$$

where $\langle \ldots \rangle_{\xi^1}$ denotes an average over the condensed patterns $\boldsymbol{\xi}^1 = (\boldsymbol{\xi}^{1,1}, \boldsymbol{\xi}^{1,2}, \ldots, \boldsymbol{\xi}^{1,s})$. In the above equations, we omitted the Γ terms corresponding to equation (35) by applying the Maxwell rule. The results coincided with those of the theoretical analysis using the replica theory [9]. For simplicity, we show the case where $s = 3, 5$. Figure 1 shows a phase diagram of numerical solutions for equations (38)–(41). The critical loading rate of a memory pattern and that of a mixed state are plotted against b. The triangular area shown in the figure represents the region where two kinds of mixed states coexist. We call this region the 'bistable region'. In this paper, a mixed pattern which is similar to $\mathrm{sgn}(\sum_{\nu=1}^{s} \boldsymbol{\xi}^{\mu,\nu})$, is defined as $\boldsymbol{\eta}^\mu$, and the other one, which is reported here for the first time, is defined as $\tilde{\boldsymbol{\eta}}^\mu$. As shown in [3], the SCSNA assumes the stability of equilibrium state. There is no free energy in the SCSNA formalism. However, we discuss the present model using the equilibrium statistical mechanics [9] since it has the energy function. From the statistical–mechanical viewpoint, the first-order transition occurs at the dashed curve in figure 1. In the region below the dashed curve, the free energy of $\boldsymbol{\eta}^\mu$ is less than that of $\tilde{\boldsymbol{\eta}}^\mu$, while the free energy of $\boldsymbol{\eta}^\mu$ is larger than that of $\tilde{\boldsymbol{\eta}}^\mu$ in the region above the dashed curve.

We will examine two typical examples in the bistable region. We consider the case where the retrieval pattern is $\boldsymbol{\xi}^{1,1}$. $\alpha_c^\eta(\alpha_c^{\tilde{\eta}})$ is defined as the critical loading rate of $\boldsymbol{\eta}^1(\tilde{\boldsymbol{\eta}}^1)$. First, we discuss the region in which $\alpha_c^\eta < \alpha_c^{\tilde{\eta}}$ as follows. Figure 2(a) shows how the overlaps between the equilibrium state and the retrieval pattern $\boldsymbol{\xi}^{1,1}$ depend on α. Two kinds of mixed states coexist when $0.015\,00 < \alpha < 0.017\,65$, as shown in figure 2($a$). From the numerical analysis, the value of the cross-talk noise variance expressed by r in equation (40) becomes larger in the order $\boldsymbol{\xi}^{1,1}$, $\boldsymbol{\eta}^1$, $\tilde{\boldsymbol{\eta}}^1$. The variation of the overlap $\frac{1}{N}\sum_{i=1}^{N} \xi_i^{1,1} \tilde{\eta}_i^1$ with α is larger than that of $\frac{1}{N}\sum_{i=1}^{N} \xi_i^{1,1} \eta_i^1$ with α, as shown in figure 2(a). Thus, we found that $\tilde{\boldsymbol{\eta}}^1$ is more influenced by the uncondensed patterns than $\boldsymbol{\eta}^1$ is. We will transform the order parameter equations given

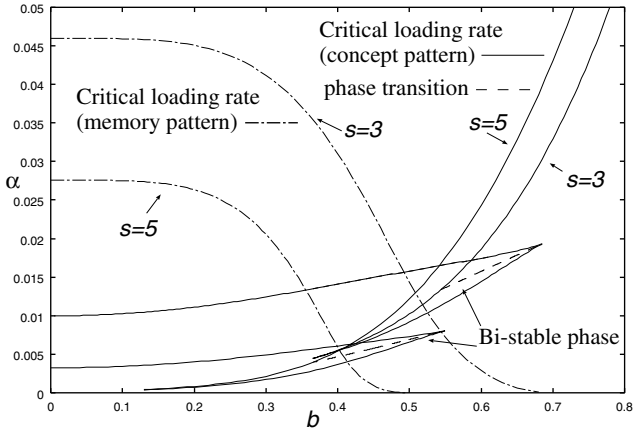

Figure 1. Variation of critical loading rate of memory pattern or mixed states with b. Two kinds of mixed states coexist in the triangular region.

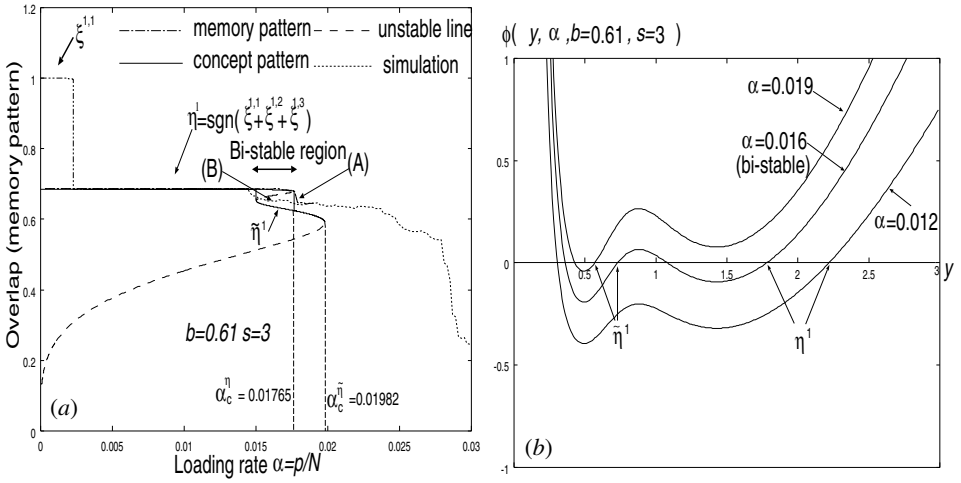

Figure 2. (*a*) Typical example ($b = 0.61$, $s = 3$) of α dependency of overlap between equilibrium state and retrieval pattern. (*b*) Solutions of transformed order parameter equation in one variable around the bistable region when $\alpha_c^{\eta} < \alpha_c^{\tilde{\eta}}$.

by equations (38)–(41) into an equation in one variable in order to qualitatively understand the nature of this bistability. Considering that $m^{1,\nu}$ is the same value for any ν in the mixed states, we can transform the order parameter equations representing the mixed states into the equation in one variable y by replacing $\frac{m^{1,\nu}}{\sqrt{2\alpha r}}$ with y^{ν} (see appendix B). Figure 2(*b*) shows the graphical solutions of the transformed equation around the bistable region. Each intersection of the y-axis and function $\Phi(y, \alpha, b = 0.61, s = 3)$ represents the corresponding solution. Figure 2(*b*) shows that two kinds of mixed states, that is, $\boldsymbol{\eta}^1$ and $\tilde{\boldsymbol{\eta}}^1$, coexist at $\alpha = 0.016$. On the other hand, there is no such bistability in the retrieval region except for the triangle region. Figure 1 shows that bistability also exists for $s = 5$. By analysing the graphical solutions of the transformed equation for various values of s, we deduced that the bistability of the mixed states exists for any value of s in model 1.

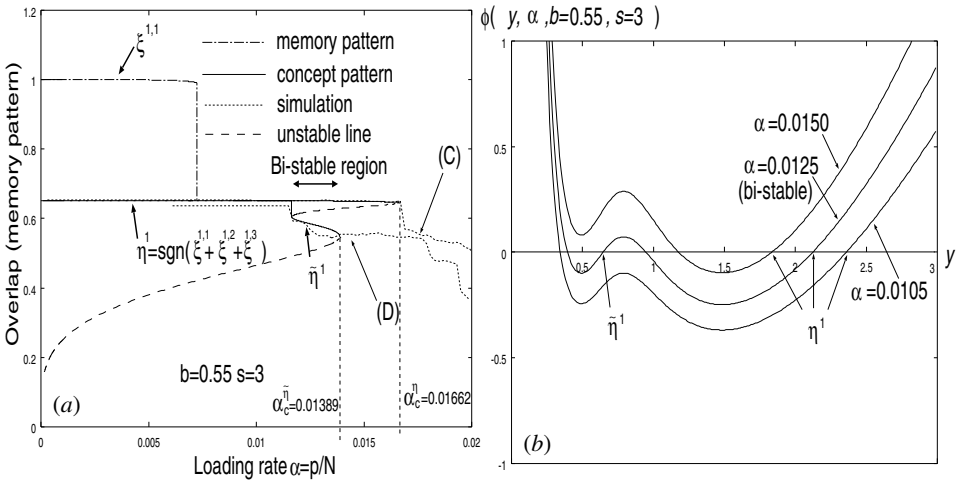

Figure 3. (*a*) Typical example ($b = 0.55$, $s = 3$) of α dependency of overlap between equilibrium state and retrieval pattern. (*b*) Solutions of transformed order parameter equation in one variable around the bistable region when $\alpha_c^{\tilde{\eta}} < \alpha_c^{\eta}$.

We performed computer simulation to confirm the bistability of mixed states, and obtained equilibrium states, which are expressed by the dotted curve (A) in figure 2(*a*), as follows. We first got an equilibrium state at $\alpha = 0.0003 < \alpha_c^{\eta}$ with the initial state set to $\mathrm{sgn}(\sum_{\nu=1}^{s} \boldsymbol{\xi}^{1,\nu})$. After that, equilibrium states for various α were obtained by gradually increasing the value of α from $\alpha = 0.0003$. Since the simulation results mostly agreed with the theoretical results, we could not distinguish between them, as shown in this figure. The dotted curve (B) in figure 2(*a*) was obtained as follows. An equilibrium state at $\alpha = 0.019$ ($\alpha_c^{\eta} = 0.017\,65 < \alpha < \alpha_c^{\tilde{\eta}} = 0.019\,82$) was obtained by gradually increasing the value of α from $\alpha = 0.0003$ in computer simulation. Using this equilibrium state, we got each equilibrium state at the corresponding α by gradually decreasing the value of α from $\alpha = 0.019$. We carried out computer simulation with $N = 10\,000$; a typical example is shown in figure 2(*a*). Figure 3(*a*) shows a typical example ($b = 0.55$, $s = 3$) of the case where $\alpha_c^{\tilde{\eta}} < \alpha_c^{\eta}$. Two kinds of mixed states coexisted when $0.011\,64 < \alpha < 0.013\,89$. The axes in this figure are the same axes as in figure 2(*a*). Figure 3(*b*) shows the graphical solutions of the transformed equation in one variable around the bistable region when $s = 3$, $b = 0.55$. This figure shows that two kinds of mixed states, that is $\tilde{\boldsymbol{\eta}}^1$ and $\boldsymbol{\eta}^1$, are stable at $\alpha = 0.0125$. The dotted curve (C) in figure 3(*a*) was obtained as follows. We first got an equilibrium state at $\alpha = 0.0003$ with the initial state set to $\mathrm{sgn}(\sum_{\nu=1}^{s} \boldsymbol{\xi}^{1,\nu})$. After that, equilibrium states for various α were obtained by gradually increasing the value of α from $\alpha = 0.0003$. Since $\boldsymbol{\eta}^1$ is always stable when $\alpha_c^{\tilde{\eta}} < \alpha < \alpha_c^{\eta}$, an equilibrium state corresponding to $\tilde{\boldsymbol{\eta}}^1$ could not be obtained by computer simulation with the initial state set to $\mathrm{sgn}(\sum_{\nu=1}^{s} \boldsymbol{\xi}^{1,\nu})$. By considering the qualitative properties of the retrieval dynamics, we carried out computer simulation in the following manner. Figure 4 shows the trajectories of temporal evolutions of overlaps of the state $\boldsymbol{x}^t$ with $\boldsymbol{\xi}^{1,1}$ or $\boldsymbol{\xi}^{1,2}$ respectively,

$$m_t^{1,\nu} = \frac{1}{N} \sum_{i=1}^{N} \xi^{1,\nu} x_i^t. \tag{42}$$

The parameters were set to $s = 3, b = 0.475, \alpha = 0.0087, N = 40\,000$. The initial states were set as $P[x_i^0 = \pm 1] = \frac{1 \pm m_0^{1,1} \xi_i^{1,1}}{2}$ with an initial overlap $m_0^{1,1}$. Theoretically, $m_t^{1,2} = m_t^{1,3}$

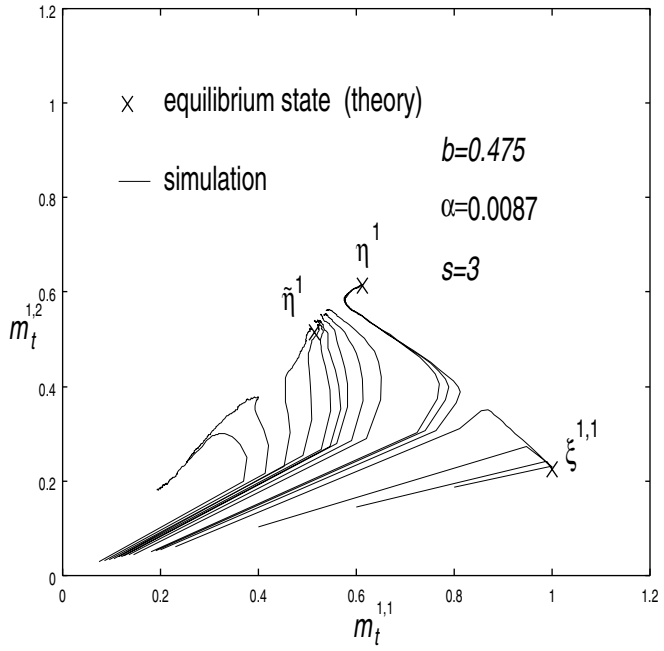

Figure 4. Retrieval processes from various initial states in computer simulation when $b = 0.475$, $s = 3$, $\alpha = 0.0087$, $N = 40\,000$.

holds over the retrieval process in this case. The crosses in figure 4 represent the theoretically obtained equilibrium states corresponding to $\tilde{\boldsymbol{\eta}}^1$, $\boldsymbol{\eta}^1$ and $\boldsymbol{\xi}^{1,1}$. Two kinds of mixed states ($\boldsymbol{\eta}^1$, $\tilde{\boldsymbol{\eta}}^1$) and the retrieval pattern ($\boldsymbol{\xi}^{1,1}$) were stable at this parameter set (figure 4). The dynamics converged to $\tilde{\boldsymbol{\eta}}^1$, $\boldsymbol{\eta}^1$, $\boldsymbol{\xi}^{1,1}$ in order as $m_0^{1,1}$ increased, while the cross-talk noise variance for the equilibrium became larger in the order, $\boldsymbol{\xi}^{1,1}$, $\boldsymbol{\eta}^1$, $\tilde{\boldsymbol{\eta}}^1$. This phenomenon occurred because the cross-talk noise variance of the initial stage of the retrieval process increased as $m_0^{1,1}$ decreased. The results showed that equilibrium states corresponding to $\tilde{\boldsymbol{\eta}}^1$ can be confirmed by computer simulation when the initial state has an appropriate overlap with the retrieval pattern. The dotted curve (D) in figure 3(a) was obtained as follows. We got an equilibrium state at $\alpha = 0.0135$ where the initial state had the appropriate overlap with the retrieval pattern ($\boldsymbol{\xi}^{1,1}$). We obtained equilibrium states for various α by increasing (decreasing) the value of α from $\alpha = 0.0135$. We carried out computer simulation with $N = 10\,000$; figure 3(a) shows a typical example. Although the results from this simulation did not always quantitatively agree with the theory owing to spurious states, we can state that the bistability of two kinds of mixed states exists in the particular region of the retrieval phase from the hysteresis shown in figures 2(a) and 3(a).

Let us present the equilibrium properties in model 2. Since the uncondensed patterns $\boldsymbol{\xi}^{\mu}(\mu = 2, 3, \ldots, \alpha N)$ of model 2 satisfy the orthogonal condition $E[\xi_i^{\mu}\xi_i^{\mu'}] = \delta_{\mu\mu'}$, the order parameter equations for this model were derived by replacing equation (40) with equation (45):

$$m^{1,\nu} = \left\langle \xi^{1,\nu} \mathrm{erf}\left(\frac{\sum_{\sigma=1}^{s} \xi^{1,\sigma} m^{1,\sigma}}{\sqrt{2\alpha r}} \right) \right\rangle_{\xi^1} \tag{43}$$

$$q = 1 \tag{44}$$

$$r = \frac{q}{(1-U)^2} \tag{45}$$

$$U = \sqrt{\frac{2}{\pi\alpha r}}\left\langle \exp\left(-\left(\sum_{\sigma=1}^{s} \frac{\xi^{1,\sigma} m^{1,\sigma}}{\sqrt{2\alpha r}} \right)^2 \right) \right\rangle_{\xi^1}. \tag{46}$$

We transformed the order parameter equations concerning the mixed states into an equation in one variable $y = \frac{m^{1,\nu}}{\sqrt{2\alpha r}}$ (see appendix B). As in model 1, we deduced that there is bistability of mixed states for an arbitrary s. These results indicated that this bistability for mixed states is not dependent on details of the structure in stored patterns.

5. Discussion

We have shown that bistability of mixed states exists in two typical models, model 1 and model 2. The reason for this bistability, which is based on the SCSNA, can be explained as follows. We introduce a model where the contribution of uncondensed patterns to the coupling J_{ij} is replaced by the spin glass type interaction, and we call this model model 3. The synaptic coupling of model 3 is defined as

$$J_{ij} = \frac{1}{N}\sum_{\nu=1}^{s} \xi_i^{1,\nu}\xi_j^{1,\nu} + \delta_{ij} \tag{47}$$

and the symmetric noise δ_{ij} is independently drawn from the following rule:

$$\delta_{ji} \sim N\left(0, \frac{\delta^2}{N}\right) \tag{48}$$

$$\delta_{ij} = \delta_{ji} \tag{49}$$

where δ is constant. We obtained the following order parameter equations for this model with the SCSNA [13]:

$$m^{1,\nu} = \int Dz \langle \xi^{1,\nu} Y \rangle_{\xi^1} \tag{50}$$

$$q = \int Dz \langle (Y)^2 \rangle_{\xi^1} \tag{51}$$

$$U = \frac{1}{\sigma}\int Dz z \langle Y \rangle_{\xi^1} \tag{52}$$

$$Y = F\left(\sum_{\sigma=1}^{s} \xi^{1,\sigma} m^{1,\sigma} + \Gamma Y + \sigma z \right) \tag{53}$$

$$\Gamma = \delta^2 U \tag{54}$$

$$Dz = \frac{dz}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \tag{55}$$

$$\sigma^2 = \delta^2 q. \tag{56}$$

This model corresponds to the Sherrington–Kirkpatrick model. In contrast to model 1 or model 2, a bistability of mixed states does not exist in model 3. Note that the analytical expression of the cross-talk noise variance expressed by equation (40) or equation (45) is explicitly expressed as a function of susceptibility U, which is a kind of a cross-talk noise-enhancement factor caused by the full-feedback nature of the model. However, the analytical expression of the cross-talk noise variance for model 3, σ in equation (56), is *not* dependent on the susceptibility U. The above results show that the bistability is due to not only the hierarchy

of stored patterns but also analytical expression of the cross-talk noise variance, which is a function of susceptibility for the model. Naturally, the bistability may exist in more complex hierarchies because the corresponding cross-talk noise variance is explicitly expressed by a function of susceptibility as shown in equation (34). From these results, we also found that the bistability of mixed states does not exist in a strongly diluted system [14].

Finally, we discuss the relation between findings concerning the retrieval process in figure 4 and the physiological experiments of face-responsive neurons in the IT cortex by Sugase *et al* [1]. Recently, Sugase *et al* have analysed the time-course of information carried by the firing of face-responsive neurons in the IT cortex, while performing a fixation task of monkey and human faces with various expressions, and simple geometrical shapes. They found that the initial transient firing correlated well with a rough categorization (e.g., face versus non-face stimuli). Their results suggest that the neuron firing pattern is initially a superposition of patterns representing different faces or expressions, but it then converges to a single pattern representing a specific face or expression. We found that the retrieval dynamics of model 1, shown in figure 4, can qualitatively replicate the temporal dynamics of face-responsive neurons as follows [15]. Initially, the network state approaches a mixed state ($\tilde{\boldsymbol{\eta}}^1$ or $\boldsymbol{\eta}^1$) that is a superposition of patterns representing different persons or expressions. After that it diverges from the mixed state, and finally converges to a single memory pattern ($\boldsymbol{\xi}^{1,1}$) representing a specific person or expression as shown in figure 4. From the above results, we expect that the present system may mimic the temporal dynamics of the face-responsive neurons [15]. Details will be discussed elsewhere.

Acknowledgments

This work was supported in part by Grants-in-aid for Scientific Research nos 09308010 and 11145229 from the Ministry of Education, Science, Sports and Culture of Japan. We are indebted to Tomoki Fukai for useful discussion.

Appendix A

The internal potential of the ith neuron h_i in the equilibrium state of equation (15) is

$$h_i = \sum_{j \neq i}^{N} J_{ij} x_j \tag{A1}$$

$$= \sum_{\rho=1}^{\hat{\alpha}N} \kappa_\rho \bar{\sigma}_i^\rho \bar{m}^\rho - \frac{1}{N} \sum_{\rho=1}^{\hat{\alpha}N} \kappa_\rho x_i . \tag{A2}$$

The output x_i can be formally expressed as

$$x_i = F\left(\sum_{\rho=1}^{\hat{\alpha}N} \kappa_\rho \bar{\sigma}_i^\rho \bar{m}^\rho - \frac{1}{N} \sum_{\rho=1}^{\hat{\alpha}N} \kappa_\rho x_i \right) \tag{A3}$$

$$= \tilde{F}\left(\sum_{\rho=1}^{\hat{\alpha}N} \kappa_\rho \bar{\sigma}_i^\rho \bar{m}^\rho \right) \tag{A4}$$

where the function $\tilde{F}(\cdot)$ is given in equation (26). Here, we assume that the equilibrium state $\boldsymbol{x}$ has nonzero overlaps with $\bar{s}$ rotated patterns $\bar{\boldsymbol{\sigma}}^\rho (\rho = 1, 2, \ldots, \bar{s})$. The residual overlap

$\bar{m}^{\rho} \sim \mathrm{O}(1/\sqrt{N})$, $(s+1 \leqslant \rho \leqslant \hat{\alpha}N)$ can be derived by using a Taylor expansion

$$\bar{m}^{\rho} = \frac{1}{N}\sum_{i=1}^{N} \bar{\sigma}_i^{\rho} \tilde{F}\left(\sum_{\rho'=1}^{\hat{\alpha}N} \kappa_{\rho'} \bar{\sigma}_i^{\rho'} \bar{m}^{\rho'} \right) \tag{A5}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \bar{\sigma}_i^{\rho} x_i^{(\rho)} + \kappa_{\rho} U \bar{m}^{\rho} \tag{A6}$$

$$= \frac{1}{N(1-\kappa_{\rho}U)}\sum_{i=1}^{N} \bar{\sigma}_i^{\rho} x_i^{(\rho)} \tag{A7}$$

where

$$x_i^{(\rho)} = \tilde{F}\left(\sum_{\rho' \neq \rho}^{\hat{\alpha}N} \kappa_{\rho'} \bar{\sigma}_i^{\rho'} \bar{m}^{\rho'} \right) \tag{A8}$$

$$x_i^{(\rho)'} = \tilde{F}'\left(\sum_{\rho' \neq \rho}^{\hat{\alpha}N} \kappa_{\rho'} \bar{\sigma}_i^{\rho'} \bar{m}^{\rho'} \right) \tag{A9}$$

$$U = \frac{1}{N}\sum_{i=1}^{N} x'^{(\rho)}_i. \tag{A10}$$

Substituting equation (A7) into equation (A2), we obtain

$$h_i = \sum_{\rho=1}^{\bar{s}} \kappa_{\rho} \bar{\sigma}_i^{\rho} \bar{m}^{\rho} + \Gamma x_i + \bar{z}_i \tag{A11}$$

where Γ is defined as

$$\Gamma = \frac{1}{N}\sum_{\rho=\bar{s}+1}^{\hat{\alpha}N} \frac{{\kappa_{\rho}}^2 U}{1-\kappa_{\rho}U} \tag{A12}$$

and $\bar{z}_i$ is the effective noise so that

$$\bar{z}_i = \frac{1}{N}\sum_{\rho=\bar{s}+1}^{\hat{\alpha}N}\sum_{j\neq i}^{N} \frac{\kappa_{\rho}}{1-\kappa_{\rho}U} \bar{\sigma}_i^{\rho} \bar{\sigma}_j^{\rho} x_j^{(\rho)}. \tag{A13}$$

Note that $\bar{z}_i$ is a summation of uncondensed patterns with $\langle \bar{z} \rangle = 0$ and $\langle \bar{z}^2 \rangle = \hat{\alpha} r$,

$$r = \frac{1}{\hat{\alpha}N}\sum_{\rho=\bar{s}+1}^{\hat{\alpha}N} \frac{\kappa_{\rho}^2}{(1-\kappa_{\rho}U)^2} \frac{1}{N}\sum_{j=1}^{N} (x_j^{(\rho)})^2. \tag{A14}$$

Let us express the sum of κ_{ρ} in terms of an integration along continuous eigenvalue $\kappa(\frac{\rho}{\hat{\alpha}N}) \equiv \kappa_{\rho}$ for $p, N \to \infty$,

$$r = q\int_0^1 \mathrm{d}u \frac{\kappa(u)^2}{(1-\kappa(u)U)^2} \tag{A15}$$

$$q = \frac{1}{N}\sum_{i=1}^{N} (x_i)^2 \tag{A16}$$

$$\Gamma = \hat{\alpha}\int_0^1 \mathrm{d}u \frac{\kappa(u)^2 U}{1-\kappa(u)U}. \tag{A17}$$

We can obtain equations (22)–(28) by replacing x_i with Y.

Appendix B

Let us transform the order parameter equations for the mixed state in model 1 into the equation in one variable. We introduce y^ν in the following equation:

$$y^\nu = \frac{m^{1,\nu}}{\sqrt{2\alpha r}}. \tag{B1}$$

Since $y^\nu = y$ ($\nu = 1, 2, 3, \ldots, s$) holds in the mixed state, we can express the order parameter equations for the mixed state by the following equation in y:

$$\begin{aligned}\Phi(y, \alpha, b, s) &= \sum_{\nu=1}^{s} \left(\frac{\lambda_\nu^2}{(w(y) - \lambda_\nu \theta(y))^2} \right) - 1 \\ &= 0 \end{aligned} \tag{B2}$$

$$\theta(y) = \sqrt{\frac{2}{\pi\alpha}} \left\langle \exp\left(- \left(\sum_{\nu}^{s} y\xi^{1,\nu} \right)^2 \right) \right\rangle_{\xi^1} \tag{B3}$$

$$w(y) = \frac{1}{\sqrt{2\alpha}\, y} \left\langle \xi^{1,\nu} \mathrm{erf}\left(\sum_{\nu}^{s} y\xi^{1,\nu} \right) \right\rangle_{\xi^1} \tag{B4}$$

where λ_ν is the νth eigenvalue of the matrix $\boldsymbol{B}$. The order parameter equations for the mixed states in model 2 can be also transformed into the equation in one variable as follows:

$$\begin{aligned}\Psi(y, \alpha, b, s) = \sqrt{2\alpha} y \left(\sqrt{\frac{2}{\pi\alpha}} \left\langle \exp\left(- \left(\sum_{\nu}^{s} y\xi^{1,\nu} \right)^2 \right) \right\rangle_{\xi^1} + 1 \right) \\ - \left\langle \xi^{1,\nu} \mathrm{erf}\left(\sum_{\nu}^{s} y\xi^{1,\nu} \right) \right\rangle_{\xi^1} = 0. \end{aligned} \tag{B5}$$

References

[1] Sugase Y, Yamane S, Ueno S and Kawano K 1999 *Nature* **400** 869
[2] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530
[3] Shiino M and Fukai T 1992 *J. Phys. A: Math. Gen.* **25** L375
[4] Amari S 1977 *Biol. Cybern.* **26** 175
[5] Griniasty M, Tsodyks M V and Amit D J 1993 *Neural Comput.* **5** 1
[6] Amit D J, Brunel N and Tsodyks M V 1994 *J. Neurosci.* **14** 6435
[7] Miyashita Y 1988 *Nature* **335** 817
[8] Parga N and Rolls E 1998 *Neural Comput.* **10** 1507
[9] Fontanari J F 1990 *J. Physique* **51** 2421
[10] Shiino M and Fukai T 1993 *Phys. Rev.* E **48** 867
[11] Aonishi T, Kurata K and Okada M 1999 *Phys. Rev. Lett.* **82** 2800
[12] Mimura K, Okada M and Kurata K 1998 *IEICE. Trans. Inf. Syst.* **E81-D** 1298
[13] Okada M, Fukai T and Shiino M 1998 *Phys. Rev.* E **57** L2095–103
[14] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167
[15] Okada M, Toya K, Kimito T and Doya K 1999 *Abstr. Soc. Neurosci.* **25** part 1-917