IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **43** (2010) 025004 (22pp) doi:10.1088/1751-8113/43/2/025004

# Statistical mechanical study of partial annealing of a neural network model

**T Uezu[1], K Abe[1], S Miyoshi[2] and M Okada[3,4]**

[1] Graduate School of Humanities and Sciences, Nara Women's University, Nara 630-8506, Japan
[2] Department of Electrical and Electronic Engineering, Faculty of Engineering Science, Kansai University, Osaka 564-8680, Japan
[3] Division of Transdisciplinary Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8561, Japan
[4] RIKEN Brain Science Institute, Saitama 351-0198, Japan

E-mail: uezu@ki-rin.phys.nara-wu.ac.jp

Received 9 March 2009, in final form 7 August 2009
Published 10 December 2009
Online at stacks.iop.org/JPhysA/43/025004

**Abstract**

We study a neural network model in which both neurons and synaptic interactions evolve in time simultaneously. The time evolution of synaptic interactions is described by a Langevin equation including a Hebbian learning term with the learning coefficient $\varepsilon$, and a bias term which is the interaction of the Hopfield model. We assume that synaptic interactions change is much slower than neurons and we study the stationary states of synaptic interactions by the replica method. We draw phase diagrams taking into account the stability of solutions, and find that the temperature region in which the Hopfield attractor is stable increases as the learning coefficient increases. Theoretical results are confirmed by the direct numerical integration of the Langevin equation. Further, we study the characteristics of the resultant synaptic interactions by partial annealing in the parameter region where the Hopfield and the mixed states exist. We find two kinds of interactions, one of which has the Hopfield attractor and the other has the mixed state attractor. Each interaction is characterized mainly by the eigenvector belonging to the largest eigenvalue of the interaction as a matrix.

PACS numbers: 87.10.−e, 05.20.−y, 84.35.+i

## 1. Introduction

We are interested in systems not only with elements but also interaction changes between elements. For example, in immune systems, the affinity, i.e. the interaction strength between immune cells (B cells) and antigens, changes by somatic hypermutation. Another example is

  

neuronal systems, in which synaptic connections change, and this is considered as learning. In these two cases, elements interact much faster than the interactions change. In fact, in the former case, the time scale that B cells recognize an antigen is about several hours, and one that somatic hypermutation would take place in about one week. In the latter case, the time scale of spiking of neurons is about several milliseconds, and of that the learning in the short-term memory is in the order of minutes. That is, the time scales differ by about $10^2$ times and $10^4$ times in the former and the latter cases, respectively.

In this paper, we study a mathematical model of neuronal systems and consider the double dynamics of neurons and synaptic interactions. As mentioned above, since the time evolution of neurons is very rapid, in the model we investigate that interactions are considered to be constant during the calculation of the average of correlations of neurons. We study the stationary states of the system by the replica method. In this study, the replica number $n$ can be any value and the $n \to 0$ limit is not used unlike the usual replica method. The case whereby the synaptic interactions do not change we refer to as 'quenched', whereas when they change together with neurons we refer to this as 'annealing'. The present case in which synaptic interactions change much slower than neurons is intermediate and is called 'partial annealing'.

Previously, there have been several studies on neural networks in which both synaptic interactions and neurons evolve in time without using the replica method, e.g. [1–4]. Among others, in [1] a rule of synaptic modification was proposed by minimizing 'free energy' and the dynamics of synaptic interactions $\{J_{ij}\}$ was adiabatically slower compared to the neuronal dynamics and deterministic. Later, several authors studied the model by introducing Gaussian white noise and the bias term [5–7]. In these papers, the replica method was used. In [8], the case in which the replica number is negative was treated. A more general situation in which synaptic interactions are divided into a hierarchy of several groups, with adiabatically separated and monotonically increasing time scale, was studied in [9].

In this paper, we study the same model as that in [5–7] introducing the coefficient of the Hebbian learning and investigate the effect of the Hebbian learning in partial annealing. We study the stable states of neurons and the resultant synaptic interactions when the system reaches the stationary state [10, 11].

The idea to introduce the coefficient of learning term and investigate the system behavior by changing it is easily applied to other systems. In fact, similar analyses are under investigation for the Mexican-hat-type interactions [11] and for the Amit model [12].

In the next section, we give the formulation of the model and study the saddle point equations, the Almeida–Thouless stability (AT stability) [13] which describes the thermodynamic stability of solutions and the phase transitions. In section 3, the results of numerical simulations are presented and are compared with theoretical results. In section 4, we study the nature of interactions generated by partial annealing. In section 5, a summary and discussion are given. In the appendix, we give the analysis of the AT stability compactly.

## 2. Formulation

### *2.1. Model*

We consider the system with $N$ neurons. We assume that a neuron takes only two states, the firing state and the rest state. Let $\sigma_i$ represent the state of $i$th neuron which takes values $\pm 1$, $\sigma_i = 1$ corresponds to the firing state and $\sigma_i = -1$ corresponds to the rest state. The $i$th neuron receives a signal from the $j$th neuron via synaptic connection. Let $J_{ij}$ be the strength

of the synaptic connection. For simplicity, we consider the symmetric synaptic connection, $J_{ij} = J_{ji}$.

We consider the situation that the network has already learned the $p$ kind of memories, which are expressed by patterns $\{\xi_i^\mu\}, i = 1, \ldots, N,\ \mu = 1, \ldots, p$. $\xi_i^\mu$ takes $\pm 1$ with the probability $p(\xi_i^\mu = \frac{1}{2}) = 1 - p(\xi_i^\mu = -\frac{1}{2}) = \frac{1}{2}$ independently. The learning rule is assumed to be the Hebbian rule and the resultant interaction $K_{ij}$ is nothing but the Hopfield model which is expressed as

$$K_{ij} = \frac{K}{\sqrt{p}} \sum_{\nu=1}^{p} \xi_i^\nu \xi_j^\nu. \tag{1}$$

We assume that $J_{ij}$ converges to $K_{ij}$ if the learning term and the external noise do not exist.

Other causes of the evolution of the interaction are learning and fluctuations in the environment around synaptic connections. The learning term is expressed by $\varepsilon\sigma_i\sigma_j$. $\varepsilon$ represents the speed of the Hebbian learning. Since we assume that the time scale of neurons is much larger than that of synaptic connections, we replace $\sigma_i\sigma_j$ by the time average $\frac{1}{L}\sum_{k=1}^{L}\sigma_i(t_k)\sigma_j(t_k)$ and further to the ensemble average $\langle\sigma_i\sigma_j\rangle_{sp}$. We assume that the ensemble average $\langle\sigma_i\sigma_j\rangle_{sp}$ is calculated by the canonical distribution with the Hamiltonian $H$ with the instantaneous values of $J_{ij}$ at time $t$:

$$\langle\sigma_i\sigma_j\rangle_{sp} \equiv \frac{1}{Z_\beta}\mathrm{Tr}_{\{\sigma_i\}}\, \mathrm{e}^{-\beta H}\sigma_i\sigma_j, \tag{2}$$

$$H(\{\sigma_i\}) = -\frac{1}{2}\sum_{i\neq j} J_{ij}\sigma_i\sigma_j, \tag{3}$$

$$Z_\beta = \mathrm{Tr}_{\{\sigma\}}\, \mathrm{e}^{-\beta H}. \tag{4}$$

Here, $\mathrm{Tr}_{\{\sigma\}}$ denotes the summation of all configurations of neurons $\{\sigma_i\}$. $Z_\beta$ is the partition function of neurons. $\beta = \frac{1}{T}$ and $T$ represents the neuronal 'temperature' of the environment around neurons. When $T \to 0$, the rule of neuron firing becomes deterministic. Not only are neurons subject to various influences from their environment but synaptic connections are as well, for example, the influence of the fluctuations of the concentration of neuronal transmitters. Thus, in the process of learning, time evolution of $J_{ij}$ is subject to external noises. We express this effect by the term $\eta_{ij}$ which is a white Gaussian random variable with the mean 0 and the following covariance:

$$\langle\eta_{ij}(t)\eta_{kl}(t')\rangle = 2\tilde{T}\delta_{ik}\delta_{jl}\delta(t - t'). \tag{5}$$

Here, $\tilde{T}$ represents the strength of the synaptic noise, and is called 'temperature' of the synaptic noise.

Thus, the evolution equation for $J_{ij}$ is expressed by the following Langevin equation [6]:

$$\begin{aligned} &\tau\frac{\mathrm{d}}{\mathrm{d}t}J_{ij} = \frac{1}{N}\varepsilon\langle\sigma_i\sigma_j\rangle_{sp} + \frac{1}{N}K_{ij} - \mu J_{ij} + \eta_{ij}(t)\sqrt{\frac{\tau}{N}}, \\ &i < j = 1, \ldots, N. \end{aligned} \tag{6}$$

The only term we did not mention is the third term $-\mu J_{ij}$, and it just represents the relaxation and is introduced so that $J_{ij}$ does not diverge. The coefficients $\frac{1}{N}$ and $\frac{1}{\sqrt{N}}$ are scaling factors so that the system has non-trivial limit as $N \to \infty$.

In the present scheme, we assume that the neuronal system and the synaptic system are in the different environments and then, in general, the synaptic temperature $\tilde{T}$ is not equal to

the neuronal temperature $T$. This is plausible because the time scale of the change of neuron states is much larger than that of the change of the synaptic connections.

By defining $\mathcal{H}$ as

$$\mathcal{H} = -\sum_{i<j} K_{ij} J_{ij} + \frac{\mu N}{2} \sum_{i<j} J_{ij}^2 - \frac{\varepsilon}{\beta} \ln Z_\beta, \tag{7}$$

the Langevin equation is rewritten as follows:

$$\tau \frac{\mathrm{d}J_{ij}}{\mathrm{d}t} = -\frac{1}{N} \frac{\partial \mathcal{H}}{\partial J_{ij}} + \eta_{ij}(t) \sqrt{\frac{\tau}{N}}. \tag{8}$$

In the stationary state of equation (8), the probability density $P(\{J_{ij}\})$ of the synaptic interactions $\{J_{ij}\}$ is given by

$$P(\{J_{ij}\}) \propto \mathrm{e}^{-\tilde{\beta}\mathcal{H}}, \tag{9}$$

$$\tilde{\beta} = \frac{1}{\tilde{T}}. \tag{10}$$

Thus, the partition function $\tilde{Z}_{\tilde{\beta}}$ of the total system is expressed by

$$\begin{aligned} \tilde{Z}_{\tilde{\beta}} &= \int \mathrm{d}\boldsymbol{J}\, \mathrm{e}^{-\tilde{\beta}\mathcal{H}} \\ &= \int \mathrm{d}\boldsymbol{J}\, Z_\beta^n\, \mathrm{e}^{-\frac{\tilde{\beta}}{2} N\mu \sum_{i<j} J_{ij}^2 + \frac{\tilde{\beta}K}{\sqrt{p}} \sum_{i<j} J_{ij} \sum_{\nu=1}^{p} \xi_i^\nu \xi_j^\nu}. \end{aligned} \tag{11}$$

Here, $\mathrm{d}\boldsymbol{J} = \prod_{i<j} \mathrm{d}J_{ij}$ and $n = \varepsilon \frac{\tilde{\beta}}{\beta}$. Now, we calculate $Z_\beta^n$ by the replica method regarding $n$ as an integer. Introducing $n$ replicas $\sigma_i^1, \sigma_i^2, \ldots, \sigma_i^n$, $Z_\beta^n$ is expressed as

$$\begin{aligned} Z_\beta^n &= \prod_{\alpha=1}^{n} \mathrm{Tr}_{\{\sigma_i^\alpha\}}\, \mathrm{e}^{-\beta H(\{\sigma_i^\alpha\})} \\ &= \mathrm{Tr}_{\{\sigma_i^\alpha\}}\, \mathrm{e}^{\beta \sum_{i<j} J_{ij} \sum_\alpha \sigma_i^\alpha \sigma_j^\alpha + \beta \sum_{\nu,i} h_\nu \xi_i^\nu \sum_\alpha \sigma_i^\alpha}. \end{aligned} \tag{12}$$

Then the integration over $\{J_{ij}\}$ is performed. We define order parameters $m_\nu^\alpha$ and $q^{\alpha\beta}$ as

$$m_\nu^\alpha = \frac{1}{N} \sum_i \sigma_i^\alpha \xi_i^\nu, \tag{13}$$

$$q_{\alpha\beta} = \frac{1}{N} \sum_i \sigma_i^\alpha \sigma_i^\beta. \tag{14}$$

Then we obtain the following expressions:

$$\tilde{Z}_{\tilde{\beta}} = \sqrt{\frac{2\pi}{\tilde{\beta}} N\mu} \int \left[ \prod_{\alpha<\beta} \frac{\mathrm{i}N}{2\pi} \mathrm{d}\hat{q}_{\alpha\beta}\, \mathrm{d}q_{\alpha\beta} \right] \left[ \prod_{\alpha,\nu} \int \frac{N}{2\pi \mathrm{i}} \mathrm{d}\hat{m}_\nu^\alpha\, \mathrm{d}m_\nu^\alpha \right] \mathrm{e}^{NG}, \tag{15}$$

$$G = G_1 + G_2 + G_3, \tag{16}$$

$$G_1 = \frac{1}{4\mu} \tilde{\beta} K^2 + \frac{\beta^2}{2\mu\tilde{\beta}} \sum_{\alpha<\beta} q_{\alpha\beta}^2 + \frac{\beta^2 n}{4\mu\tilde{\beta}} + \frac{\beta K}{2\mu\sqrt{p}} \sum_{\alpha,\nu} \left(m_\nu^\alpha\right)^2 + \beta \sum_{\alpha,\nu} h_\nu m_\nu^\alpha, \tag{17}$$

$$G_2 = -\sum_{\alpha<\beta} \hat{q}_{\alpha\beta} q_{\alpha\beta} + \sum_{\alpha,\nu} \hat{m}_\nu^\alpha m_\nu^\alpha, \tag{18}$$

$$G_3 = \frac{1}{N}\sum_i \ln\left\{\mathrm{Tr}_{\{\sigma_i^\alpha\}}\, \mathrm{e}^{\sum_{\alpha<\beta}\hat{q}_{\alpha\beta}\sigma_i^\alpha\sigma_i^\beta - \sum_{\alpha,\nu}\hat{m}_\nu^\alpha\sigma_i^\alpha\xi_i^\nu}\right\} \tag{19}$$

$$= \left[\ln\left\{\mathrm{Tr}_{\{\sigma^\alpha\}}\, \mathrm{e}^{\sum_{\alpha<\beta}\hat{q}_{\alpha\beta}\sigma^\alpha\sigma^\beta - \sum_{\alpha,\nu}\hat{m}_\nu^\alpha\sigma^\alpha\xi^\nu}\right\}\right]. \tag{20}$$

Since we consider the finite number of patterns $p$ and take $N \to \infty$ limit in this paper, $N \gg 2^p$ holds. Therefore, in the first equation of the expression $G_3$, $\frac{1}{N}\sum_i$ can be replaced by the average over $\{\xi\}$. We denote this average by $[\cdots]$ and get the second equality.

Now, we assume the replica symmetry (RS) as

$$\begin{aligned} q_{\alpha\beta} &= q, \\ m_\nu^\alpha &= m_\nu, \\ \hat{q}_{\alpha\beta} &= \hat{q}, \\ \hat{m}_\nu^\alpha &= \hat{m}_\nu. \end{aligned} \tag{21}$$

Let $G_{i,\mathrm{RS}}$ be $G_i$ evaluated at the RS solution. Then we obtain the following expressions for $G_{1,\mathrm{RS}}$, $G_{2,\mathrm{RS}}$ and $G_{3,\mathrm{RS}}$:

$$G_{1,\mathrm{RS}} = \frac{1}{4\mu}\tilde{\beta}K^2 + \frac{\beta^2}{\mu\tilde{\beta}}\frac{n(n-1)}{4}q^2 + \frac{\beta^2 n}{4\mu\tilde{\beta}} + \frac{\beta K}{2\mu\sqrt{p}}n\sum_\nu m_\nu^2 + \beta n\sum_\nu h_\nu m_\nu, \tag{22}$$

$$G_{2,\mathrm{RS}} = -\frac{n(n-1)}{2}\hat{q}q + n\sum_\nu \hat{m}_\nu m_\nu, \tag{23}$$

$$\begin{aligned} G_{3,\mathrm{RS}} &= -\hat{q}\frac{1}{2}n + \left[\ln\int Dx\left\{\cosh\left(\sqrt{\hat{q}}x - \sum_\nu \hat{m}_\nu\xi^\nu\right)\right\}^n\right] + n\ln 2, \\ Dx &= \frac{dx}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}. \end{aligned} \tag{24}$$

### 2.2. Saddle point equations (SPEs)

Defining $G_{\mathrm{RS}} = G_{1,\mathrm{RS}} + G_{2,\mathrm{RS}} + G_{3,\mathrm{RS}}$, we obtain the following saddle point equations, where we put $h_\nu = 0$ and define $J = \frac{K}{\mu\sqrt{p}}$ and $\kappa = \frac{\beta^2}{\mu\tilde{\beta}}$:

$$\frac{\partial G_{\mathrm{RS}}}{\partial q} = 0 : \hat{q} = \kappa q, \tag{25}$$

$$\frac{\partial G_{\mathrm{RS}}}{\partial \hat{q}} = 0 : q = \left[\int Dx\cosh^n\Xi\tanh^2\Xi\left\{\int Dx\cosh^n\Xi\right\}^{-1}\right], \tag{26}$$

$$\frac{\partial G_{\mathrm{RS}}}{\partial m_\nu} = 0 : \hat{m}_\nu = -\beta J m_\nu, \tag{27}$$

$$\frac{\partial G_{\mathrm{RS}}}{\partial \hat{m}_\nu} = 0 : m_\nu = \left[\xi^\nu\int Dx\cosh^n\Xi\tanh\Xi\left\{\int Dx\cosh^n\Xi\right\}^{-1}\right], \tag{28}$$

$$\Xi = \sqrt{\kappa q}x + \beta\sum_\nu J m_\nu\xi^\nu. \tag{29}$$

In this paper, we study the case of $p = 3$ and abbreviate the SPEs as

$$q \equiv \varphi(q, m_1, m_2, m_3), \tag{30}$$

$$m_\nu \equiv \psi_\nu(q, m_1, m_2, m_3)\ (\nu = 1, 2, 3). \tag{31}$$

We find the following solutions of the SPEs:

- paramagnetic solution ($P$): $q = 0, m_\mu = 0$,
- spin glass solution ($SG$): $q > 0, m_\mu = 0$,
- Hopfield attractor ($H$): $q > 0, m_1 \neq 0, m_2 = m_3 = 0$,
- mixed state with two patterns: $q > 0, m_1 = m_2 \neq 0, m_3 = 0$,
- mixed state with three patterns ($M$): $q > 0, m_1 = m_2 = m_3 \neq 0$.

In this paper, we analyze the Hopfield attractor, the mixed state with three patterns and the spin glass state. We do not consider the mixed state with two patterns because it is expected to be unstable, so hereafter, we refer to the mixed state with three patterns.

### *2.3. AT stability*

We study the AT stability of the RS solution [13]. The condition of the stability is that the free energy increases when order parameters deviate from those at the RS solution. The free energy per neuron $\tilde{f}_{\tilde{\beta}}$ is given by

$$\tilde{f}_{\tilde{\beta}} = -\frac{1}{N}\frac{1}{\tilde{\beta}} \ln \tilde{Z}_{\tilde{\beta}} = -\frac{G}{\tilde{\beta}}. \tag{32}$$

We define small deviations from the RS solution by

$$m_\nu^\alpha = m_\nu + \epsilon_\nu^\alpha,$$
$$q_{\alpha\beta} = q + \eta^{\alpha\beta},$$

and expand $G$ up to the second order of deviations. Then $G$ is expressed as

$$G = G_{\mathrm{RS}} + \frac{1}{2}\sum_{(\alpha\nu)(\beta\mu)} \mathcal{G}_{(\alpha\nu)(\beta\mu)}\epsilon_\nu^\alpha \epsilon_\mu^\beta + \frac{1}{2}\sum_{(\alpha\nu)(\beta\gamma)} \mathcal{G}_{(\alpha\nu)(\beta\gamma)}\epsilon_\nu^\alpha \eta^{\beta\gamma} + \frac{1}{2}\sum_{(\alpha\beta)(\gamma\delta)} \mathcal{G}_{(\alpha\beta)(\gamma\delta)}\eta^{\alpha\beta}\eta^{\gamma\delta}, \tag{33}$$

$G_{\mathrm{RS}} = G\,|_{q_{\alpha\beta}=q, m_\nu^\alpha = m_\nu}$.

$\mathcal{G}$ is called the Hessian. The stability condition of the RS solution is that all eigenvalues of $\mathcal{G}$ are negative. We calculated all eigenvalues for the Hopfield attractor, the mixed state and spin glass state. There exist seven different kinds of eigenvalues for the Hopfield attractor and the mixed state, and five different kinds of eigenvalues for the spin glass state. Details are given in the appendix.

### *2.4. Phase transition*

In this subsection, we study the phase transition. The second-order phase transition temperatures are determined by the following relations:

$$T_{P\to H}^{\mathrm{2nd}} = T_{P\to M}^{\mathrm{2nd}} = \frac{K}{\mu\sqrt{p}} \equiv J, \tag{34}$$

$$T_{P\to SG}^{\mathrm{2nd}} = \frac{1}{\sqrt{\mu\tilde{\beta}}} = \sqrt{\frac{\tilde{T}}{\mu}}, \tag{35}$$

$$T_{SG\to H}^{\mathrm{2nd}} = J\left\{\left(\varepsilon\frac{T_{SG\to H}^{\mathrm{2nd}}}{\tilde{T}} - 1\right)q + 1\right\}, \tag{36}$$

$$T_{SG\to M}^{\mathrm{2nd}} = T_{SG\to H}^{\mathrm{2nd}}. \tag{37}$$

Since the second-order phase transition temperature from the paramagnetic state to the Hopfield attractor, $T^{\rm 2nd}_{P\to H}$, and that from the paramagnetic state to the mixed state, $T^{\rm 2nd}_{P\to M}$, are equal, we denote them by $T^{\rm 2nd}_{P\to HM}$. Here $HM$ implies both the Hopfield attractor and the mixed state. Similarly, since $T^{\rm 2nd}_{SG\to H}$ and $T^{\rm 2nd}_{SG\to M}$ are equal, we denote them by $T^{\rm 2nd}_{SG\to HM}$. Next, we study the first-order phase transition. In this case, a new phase appears suddenly irrespective of the old phase from which the transition takes place. Thus, we consider the following three phase transitions and obtain the equations to determine the phase transition temperatures.

(i) Transition to the Hopfield attractor

$$q = \varphi(q, m_1, 0, 0), \tag{38}$$

$$m_1 = \psi_1(q, m_1, 0, 0), \tag{39}$$

$$\left(1 - \frac{\partial\varphi}{\partial q}\right)\left(1 - \frac{\partial\psi_1}{\partial m_1}\right) - \frac{\partial\varphi}{\partial m_1}\frac{\partial\psi_1}{\partial q} = 0. \tag{40}$$

(ii) Transition to the mixed state

$$q = \varphi(q, m, m, m), \tag{41}$$

$$m = \psi_1(q, m, m, m), \tag{42}$$

$$\left(1 - \frac{\partial\varphi}{\partial q}\right)\left(1 - \frac{\partial\psi_1}{\partial m}\right) - \frac{\partial\varphi}{\partial m}\frac{\partial\psi_1}{\partial q} = 0. \tag{43}$$

Derivatives in the above equations are calculated as

$$\begin{aligned}\frac{\partial\psi_1}{\partial m} &= \frac{\partial}{\partial m}\psi_1(q, m, m, m)\\ &= \frac{\partial\psi_1}{\partial m_1} + \frac{\partial\psi_1}{\partial m_2} + \frac{\partial\psi_1}{\partial m_3} = \frac{\partial\psi_1}{\partial m_1} + 2\frac{\partial\psi_1}{\partial m_2},\\ \frac{\partial\varphi}{\partial m} &= \frac{\partial\varphi}{\partial m_1} + \frac{\partial\varphi}{\partial m_2} + \frac{\partial\varphi}{\partial m_3} = 3\frac{\partial\varphi}{\partial m_1}.\end{aligned}$$

(iii) Transition to $SG$

$$q = \varphi(q, 0, 0, 0), \tag{44}$$

$$\frac{\partial\varphi}{\partial q} = 1. \tag{45}$$

We solve the above equations numerically and determine the phase transition temperatures.

We have interest in how the emerging attractor changes as the parameters $\varepsilon$, $T$ and $\tilde{T}$ change. In order to see this, taking into account the AT stability, we draw the phase diagram in the $T$–$\varepsilon$ plane for $\tilde{T} = 0.1$ and 0.4 in figure 1. Hereafter, we fix $K = 1$ and $\mu = 1$.

We list the common and different features in the results for $\tilde{T} = 0.1$ and 0.4 below.

**Common features**

(i) When temperature is decreased from high temperature, the phase transition which takes place firstly is the second-order one for small values of $\varepsilon$ and is the first-order one for large values of $\varepsilon$.

(ii) When temperature is decreased from high temperature, the Hopfield attractor appears from the paramagnetic phase by the first-order phase transition at $T = T_H^{\rm 1st}$ for a large value of $\varepsilon$.

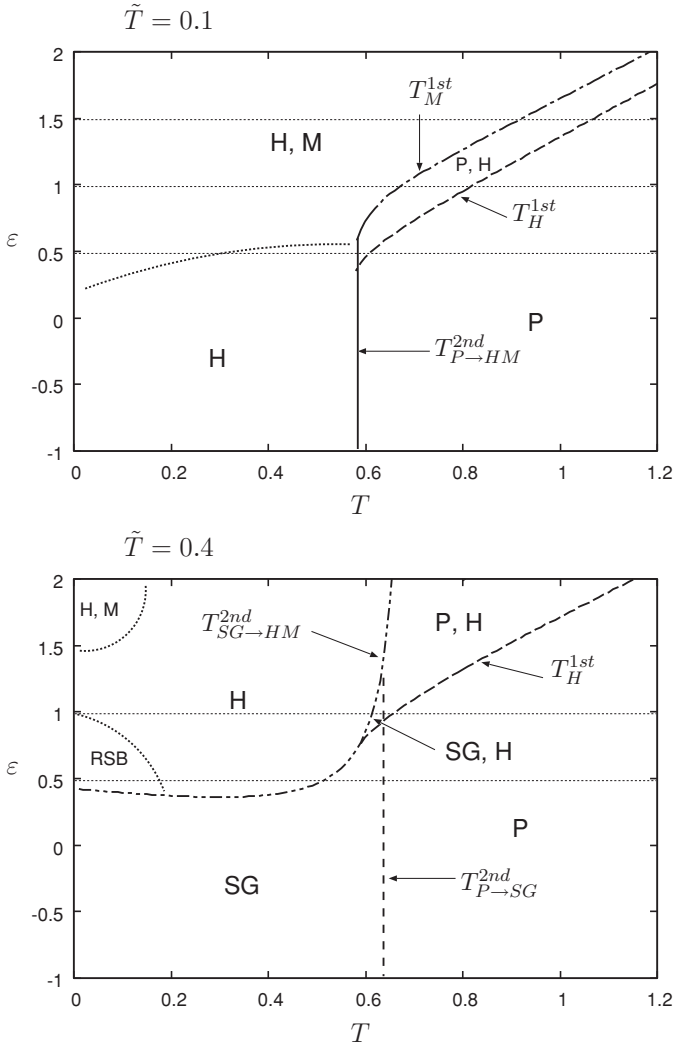


**Figure 1.** Phase diagram in the $T$–$\varepsilon$ space. $K = 1.0$, $\mu = 1.0$, solid curve: $T_{P\to HM}^{2\mathrm{nd}}$, short-dashed curve: $T_{P\to SG}^{2\mathrm{nd}}$, dashed–dotted–dotted curve: $T_{SG\to HM}^{2\mathrm{nd}}$, long-dashed curve: $T_H^{1\mathrm{st}}$, dashed–dotted curve: $T_M^{1\mathrm{st}}$. Dotted horizontal lines indicate the parameter where simulations were performed. Dotted curves denote phase boundaries which are not calculated theoretically but estimated by numerical data available. RSB denotes the replica symmetry breaking state of the Hopfield attractor.

(iii) As $\varepsilon$ is increased, the first-order phase transition temperature from the paramagnetic phase to the Hopfield attractor, $T_H^{1\mathrm{st}}$, increases and then the temperature region where the Hopfield attractor is stable increases.

**Different features**

(i) The spin glass phase does not exit for $\tilde{T} = 0.1$, whereas it exists for $\tilde{T} = 0.4$.

(ii) For $\tilde{T} = 0.4$, the replica symmetry breaking of the Hopfield attractor takes place when temperature is low.

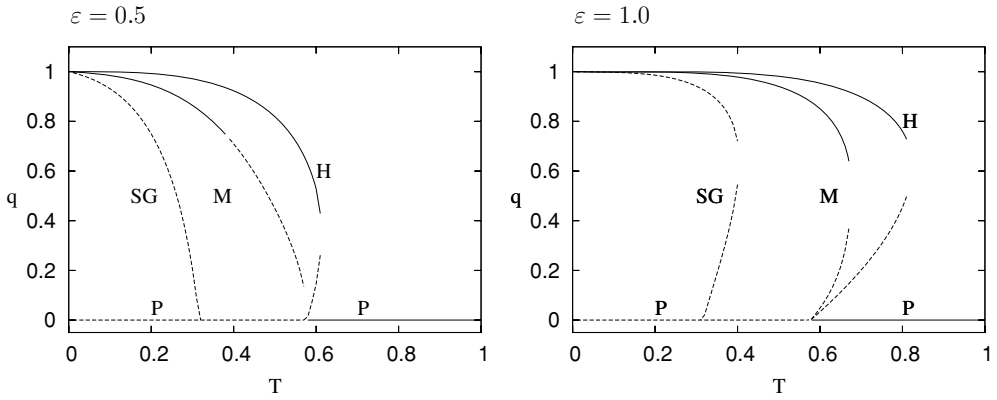


**Figure 2.** Temperature dependence of $q$. $\tilde{T} = 0.1$, $K = 1.0$ and $\mu = 1.0$. Solid curve: stable solution. Dotted curve: unstable solution.

The different features can be understood theoretically as follows.

(i) From equations (34) and (35), we note the following. $T^{2\text{nd}}_{P\to H} = J$ and $T^{2\text{nd}}_{P\to SG} = \sqrt{\frac{\tilde{T}}{\mu}}$ do not depend on $\varepsilon$. Since the former does not depend on $\tilde{T}$ neither, when $\tilde{T}$ is small, $T^{2\text{nd}}_{P\to H} > T^{2\text{nd}}_{P\to SG}$ and the Hopfield attractor appears firstly, whereas when $\tilde{T}$ is large, $T^{2\text{nd}}_{P\to H} < T^{2\text{nd}}_{P\to SG}$ and the spin glass phase appears firstly. Physically, this is interpreted as follows. Since $\tilde{T}$ is the strength of the synaptic noise, if $\varepsilon = 0$, then $n = 0$, and the present model becomes the Sherrington Kirkpatrick model (SK model) of spin glass in which the average of the interaction depends on the suffix $ij$. Since the variance of $J_{ij}$ is equal to $\frac{\tilde{T}}{N\mu}$, when $\tilde{T}$ is large, the synaptic noise becomes large, and the system tends to be spin glass. That is, the global order cannot exist but the local order is formed.

(ii) In the SK model, the replica symmetry of the spin glass solution, which exists at low temperature, breaks infinitely many times when the average of $J_{ij}$ is 0. Although the average of $J_{ij}$ is not 0 and $\varepsilon$ is not 0 in the present model, similar behavior to the SK model is expected.

Now, let us see the change of order parameter as the temperature is decreased from high temperature. Since we would like to study the phase change including the spin glass phase, we investigate the temperature dependence of $q$. Later, we investigate $m$.

We display the temperature dependence of $q$ in figure 2 for $\tilde{T} = 0.1$ and figure 3 for $\tilde{T} = 0.4$. In these figures, AT stability is also taken into account. The solid curves and dashed curves correspond to AT stable and AT unstable solutions, respectively. We can see how the first-order and second-order phase transitions take place, and the stability of solutions changes in various manners. For example, let us see the left panel of figure 2. This corresponds to the line of $\varepsilon = 0.5$ in the upper panel of figure 1. As temperature is decreased, the stable solution changes as

para $\to$ Hopfield attractor and para $\to$ Hopfield attractor $\to$
Hopfield attractor and mixed state.

Next, let us see the left panel of figure 3. This corresponds to the line of $\varepsilon = 0.5$ in the lower panel of figure 1. As temperature is decreased, the stable solution changes as

para $\to$ spin glass $\to$ Hopfield attractor $\to$
Hopfield attractor (RSB).

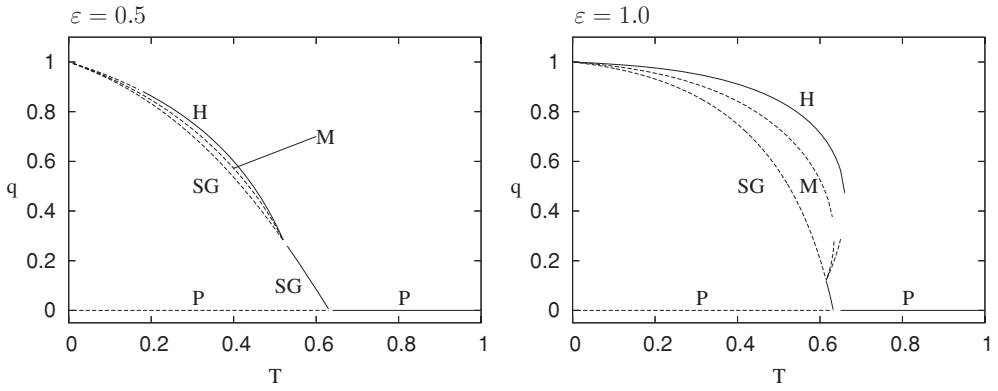


**Figure 3.** Temperature dependence of $q$. $\tilde{T} = 0.4$, $K = 1.0$ and $\mu = 1.0$. Solid curve: stable solution. Dotted curve: unstable solution.

**Table 1.** $\tilde{T} = 0.1$, $K = 1$, $\mu = 1.0$. Phase transition temperature of the Hopfield attractor and the mixed state, and the ratio $r$ of the stable temperature region for the mixed state to that for the Hopfield attractor by taking into account the AT stability for several values of $\varepsilon$.

| | The Hopfield attractor | The mixed state | $r = \frac{\text{stable region for the mixed state}}{\text{stable region for the Hopfield attractor}}$ |
|---|---|---|---|
| $\varepsilon = 0$ | 0.58 | 0.27 | 0.47 |
| $\varepsilon = 0.5$ | 0.61 | 0.38 | 0.62 |
| $\varepsilon = 1.0$ | 0.83 | 0.68 | 0.82 |
| $\varepsilon = 1.5$ | 1.07 | 0.92 | 0.86 |

The last transition is due to the loss of the AT stability of the RS solution of the Hopfield attractor, that is, the replica symmetry breaking (RSB) solution appears.

As seen from these figures, various phases appear depending on the values of $T, \tilde{T}, \varepsilon$.

In table 1, we show the phase transition temperature of the Hopfield attractor, $T_c^H$, that of the mixed state, $T_c^M$, and the ratio $r$ of the stable temperature region for the mixed state to that for the Hopfield attractor by taking into account the AT stability for several values of $\varepsilon$ and $\tilde{T} = 0.1$. We note that $T_c^H$, $T_c^M$ and $r$ increase as $\varepsilon$ increases from 0.

## 3. Simulations

We set $\tau = 1$ without loss of generality, and $K$ and $\mu$ are set to 1 for simplicity. We performed direct integrations of the Langevin equation (6) changing parameters $T, \tilde{T}$ and $\varepsilon$. We used the Euler method with the time increment $\Delta t = 0.1$. We adopted the following procedure according to [6]. When $N$ times of update by the Monte Carlo method are tried, we call it 1 Monte Carlo step and denote it by 1[MCS].

(i) Set an initial state of interactions $\boldsymbol{J}(0)$.
(ii) Update neurons $\boldsymbol{\sigma}$ by $R_1$ [MCS].
(iii) Calculate $\langle \sigma_i \sigma_j \rangle$ during the $R_2$ [MCS] update of neurons.
(iv) Update $\boldsymbol{J}$ by the Euler method.

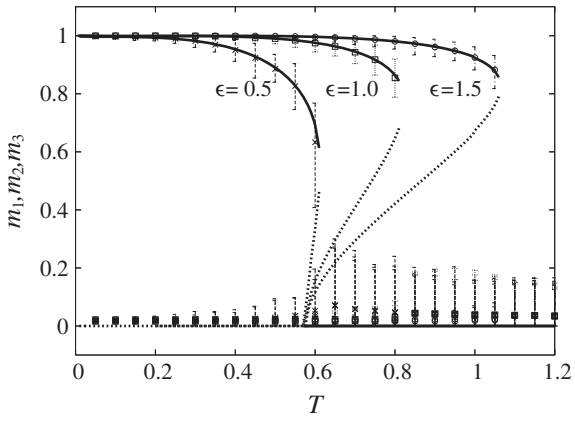


**Figure 4.** Temperature dependence of $m_1, m_2, m_3$ for the Hopfield attractor. $\tilde{T} = 0.1$, $K = 1.0$ and $\mu = 1.0$. Curves are theoretical results. Solid curves are stable and dotted curves are unstable. Symbols are simulation results for $N = 1000$.

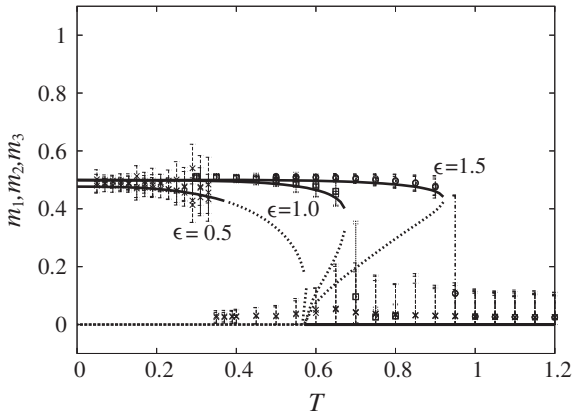


**Figure 5.** Temperature dependence of $m_1, m_2, m_3$ for the mixed state. $\tilde{T} = 0.1$, $K = 1.0$ and $\mu = 1.0$. Curves are theoretical results. Solid curves are stable and dotted curves are unstable. Symbols are simulation results for $N = 2000$.

(v) Repeat two to four $R_3$ times.
(vi) Calculate averages of physical quantities during the $R_4$ repetitions of two to four.

In this procedure, the total number of updates of neurons is $(R_1 + R_2)(R_3 + R_4)$ [MCS]. As $R$s, we took $R_1 = R_2 = R_3 = R_4 = 500$.

We show the numerical results for $m_1, m_2, m_3$ together with theoretical ones for the Hopfield attractor in figure 4 and for the mixed state in figure 5.

We chose $\tilde{T} = 0.1$ because the parameter region in $(T, \varepsilon)$ where the Hopfield attractor and the mixed state appear is wide as seen from figure 1. We chose $\varepsilon = 0.5, 1.0, 1.5$. $\varepsilon = 1$ is chosen because this value corresponds to the situation that the order of the magnitude of the learning term $\langle \sigma_i \sigma_j \rangle_{sp}$ and the term $K_{ij}$ are equal. Two other values are chosen because we focus on the case of positive values of $\varepsilon$ and we wanted to choose a larger and a smaller value of $\varepsilon$ compared with $\varepsilon = 1$.

We chose the system size $N = 1000$ in the simulation for the Hopfield attractor and $N = 2000$ in that for the mixed state, and confirmed that these values are appropriate for getting results with sufficient accuracy in reasonable machine time.

From the theoretical results, we expect the following properties.

(i) The upper branches of the solutions are stable and their lower branches are unstable.
(ii) As $\varepsilon$ is increased, the temperature region in which the Hopfield attractor and/or the mixed state exist becomes wider.
(iii) The Hopfield attractor and the mixed state appear by the first-order phase transition at $T = T_H^{1st}$ and $T = T_M^{1st}$, respectively.
(iv) There is one exception of the above statements. That is, in the case of the mixed state at $\varepsilon = 0.5$, the RS solution is unstable above $T = 0.38$, which is less than $T = T_M^{1st}$.

All of the these theoretical predictions are confirmed by the simulations. The agreement between simulations and the stable upper branch solutions by theory is good, except for two cases in the simulations of the mixed state. One is just above the critical temperature for $\varepsilon = 1.0$ and $\varepsilon = 1.5$. In the paramagnetic phase, fluctuations are very large. The other is around the temperature where the AT instability takes place for $\varepsilon = 0.5$. In this case, the phase transition takes place at the temperature which is lower than the theoretical prediction. Results in both cases are considered to be finite-size effects.

## 4. On the nature of interactions generated by partial annealing

In this section, we numerically study the interaction $\{J_{ij}\}$ which appears after partial annealing. There are two types of such interactions. One is $\left\{J_{ij}^H\right\}$ in which the Hopfield attractor ($m_1 > 0, m_2 \simeq 0, m_3 \simeq 0$) appears during partial annealing, and the other is $\left\{J_{ij}^M\right\}$ in which the mixed state ($m_1 \simeq m_2 \simeq m_3 > 0$) appears. Using these two interactions, we performed Monte Carlo simulations taking the following initial conditions: three kinds of Hopfield attractor like initial conditions, $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, and eight kinds of mixed state like initial conditions, $(\pm 1/2, \pm 1/2, \pm 1/2)$, and random configurations. We describe the results obtained at $\tilde{T} = 0.1, \varepsilon = 1.0$. The results at other parameters are the same as those at the present parameters.

In figure 6, we display the time series of $m_1, m_2, m_3$ at $T = 0.4$. We found that when the interaction is $\left\{J_{ij}^H\right\}$, the neuron system converges to the Hopfield attractor, ($m_1 > 0, m_2 \simeq m_3 \simeq 0$) or ($m_1 < 0, m_2 \simeq m_3 \simeq 0$), whereas when the interaction is $\left\{J_{ij}^M\right\}$, it converges to the mixed state, ($m_1 \simeq m_2 \simeq m_3 > 0$) or ($m_1 \simeq m_2 \simeq m_3 < 0$). That is, irrespective of initial conditions, the neuron system converges to the attractor which appeared in the process of partial annealing, or to the reversed attractor whose elements $m_\mu$ have opposite sign to the attractor.

Next, we study the temperature dependence of the order parameters $m_1, m_2$ and $m_3$ which are estimated after the system becomes stationary. At each temperature $T$, we use the interaction $\left\{J_{ij}^H\right\}$ or $\left\{J_{ij}^M\right\}$, and perform the Monte Carlo simulation. When the interaction $\left\{J_{ij}^H\right\}$ or $\left\{J_{ij}^M\right\}$ is used, the initial condition is set to the mixed state or to the Hopfield attractor, respectively. In figure 7, we display the theoretical and simulation results for the temperature dependence of $m_1, m_2$ and $m_3$. As seen from these figures, the resultant attractor by the Monte Carlo simulation is the Hopfield attractor for $\left\{J_{ij}^H\right\}$ and the mixed state for $\left\{J_{ij}^M\right\}$, respectively. In the right panel of figure 7, the Hopfield attractor appears at $T = 0.7, 0.75$ and $0.8$ because the mixed state is unstable at these temperatures. The theoretical and numerical results agree very well.

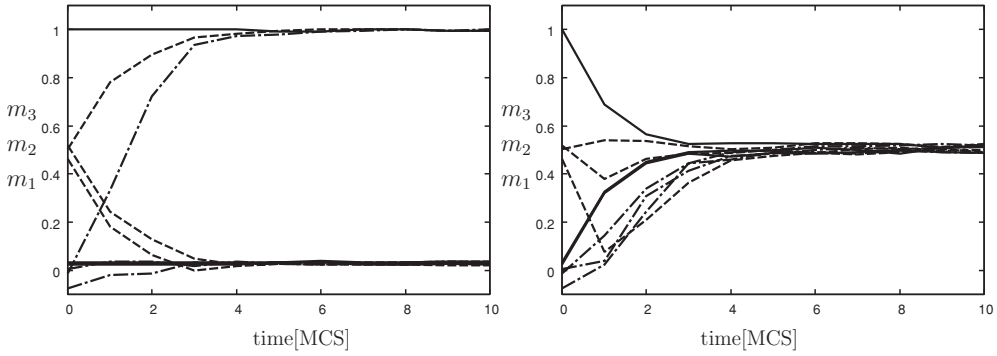


**Figure 6.** Time series of $m_1, m_2, m_3$. Numerical results ($N = 500$). $K = 1, \mu = 1, \tau = 1$. $\tilde{T} = 0.1, \varepsilon = 1.0, T = 0.4$. Left panel: interaction $\{J_{ij}^H\}$, in which the Hopfield attractor appears during partial annealing, is used. Right panel: interaction $\{J_{ij}^M\}$, in which the mixed state appears during partial annealing, is used. Initial state. Solid curve: $m_1 = 1, m_2 = m_3 = 0$; dashed curve: $m_1 \simeq m_2 \simeq m_3 \simeq 0.5$; dashed–dotted curve: random configuration.

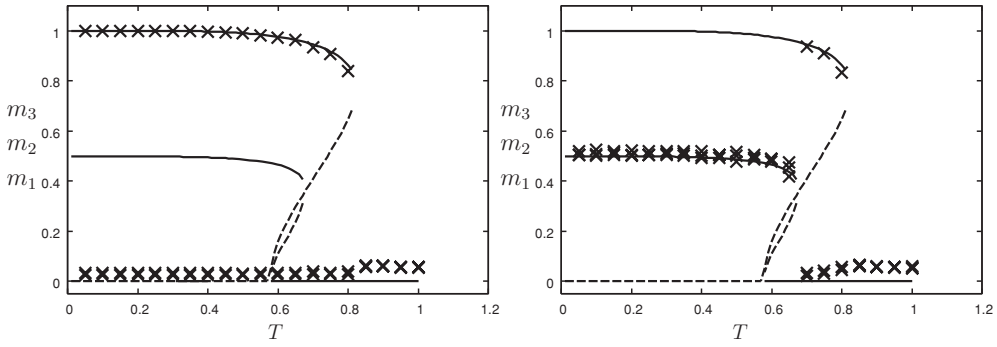


**Figure 7.** Temperature dependence of $m_1, m_2, m_3$. $\tilde{T} = 0.1, \varepsilon = 1.0$. Curves are theoretical results of partial annealing and symbols are simulation results ($N = 500$). Left panel: at each temperature, interaction $\{J_{ij}^H\}$ in which the Hopfield attractor appears during partial annealing is used. Initial state is the mixed state. Right panel: at each temperature, interaction $\{J_{ij}^M\}$ in which the mixed state appears during partial annealing is used. Initial state is the Hopfield attractor.

Now, we study interactions $\left\{J_{ij}^H\right\}$ and $\left\{J_{ij}^H\right\}$ in more detail. We regard $\left\{J_{ij}^H\right\}$, $\left\{J_{ij}^H\right\}$ and $\{K_{ij}\}$ as ${}_nC_2$ dimensional vectors and denote them by $\boldsymbol{J}^H$, $\boldsymbol{J}^M$ and $\boldsymbol{K}$, respectively. In figure 8, we display the temperature dependence of angles among these vectors. From this figure, we note that $\boldsymbol{J}^H$ and $\boldsymbol{J}^M$ have almost the same direction and they are almost perpendicular to $\boldsymbol{K}$. Also, their norms are almost same and about 5.3. In order to clarify the difference between $\left\{J_{ij}^H\right\}$ and $\left\{J_{ij}^M\right\}$, we investigate the eigenvalues and eigenvectors. In figure 9, we display the spectra of eigenvalues at $T = 0.4$. From this, we note that the spectra for $\left\{J_{ij}^H\right\}$ and $\left\{J_{ij}^M\right\}$ are almost the same except for the largest eigenvalues. We calculate direction cosines and angles between eigenvectors and pattern vectors, $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \boldsymbol{\xi}^3$ and $\boldsymbol{\xi}^{\mathrm{mix}}$. Here, the $i$th component of $\boldsymbol{\xi}^{\mathrm{mix}}$ is given by

$$\xi_i^{\mathrm{mix}} = \mathrm{sgn}\left(\xi_i^1 + \xi_i^2 + \xi_i^3\right),$$

where $\mathrm{sgn}(x) = 1$ for $x \geqslant 0$ and $-1$ otherwise. As a result, we found that the eigenvectors belonging to the three largest eigenvalues have large direction cosines with some of pattern vectors, whereas the direction cosines for the other eigenvectors are almost

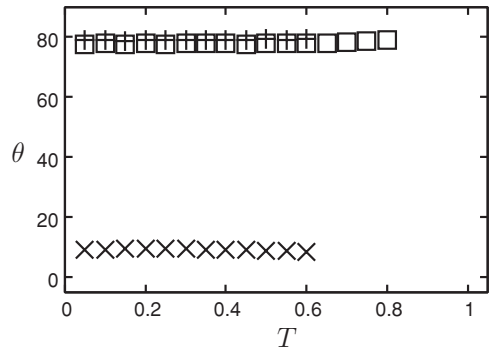


**Figure 8.** $\tilde{T} = 0.1, \varepsilon = 1.0, N = 500$. Temperature dependence of angle (degree) between interaction vectors. $\times$ : $(\hat{\boldsymbol{J}}^H, \hat{\boldsymbol{J}}^M)$, square: $(\hat{\boldsymbol{J}}^H, \hat{\boldsymbol{K}})$, + : $(\hat{\boldsymbol{J}}^M, \hat{\boldsymbol{K}})$.

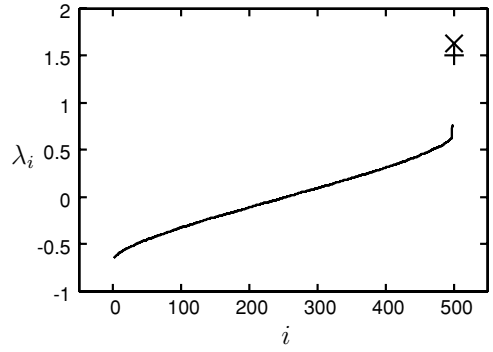


**Figure 9.** The spectra of eigenvalues of $\{J_{ij}^H\}$ and $\{J_{ij}^M\}$. $\tilde{T} = 0.1, \varepsilon = 1.0, T = 0.4, N = 500$. $\times$: largest eigenvalue of $\{J_{ij}^H\}$, +: largest eigenvalue of $\{J_{ij}^M\}$. Solid curves are eigenvalues spectra besides the largest one and are indistinguishable in $\{J_{ij}^H\}$ and $\{J_{ij}^M\}$.

0. Let $\lambda_{\max}, \lambda_{2\mathrm{nd}}, \lambda_{3\mathrm{rd}}, \lambda_{4\mathrm{th}}$ be the four largest eigenvalues and $|\lambda_{\max}\rangle, |\lambda_{2\mathrm{nd}}\rangle, |\lambda_{3\mathrm{rd}}\rangle, |\lambda_{4\mathrm{th}}\rangle$ be the normalized eigenvectors belonging to them, respectively. Since $\{J_{ij}\}$ is symmetric, all eigenvectors are orthogonal. Then $J^H \equiv \left\{J_{ij}^H\right\}$ and $J^M \equiv \left\{J_{ij}^M\right\}$ are expressed as

$$J^H = \lambda_{\max}^H \left|\lambda_{\max}^H\right\rangle\left\langle\lambda_{\max}^H\right| + \lambda_{2\mathrm{nd}}^H \left|\lambda_{2\mathrm{nd}}^H\right\rangle\left\langle\lambda_{2\mathrm{nd}}^H\right| + \lambda_{3\mathrm{rd}}^H \left|\lambda_{3\mathrm{rd}}^H\right\rangle\left\langle\lambda_{3\mathrm{rd}}^H\right| + \cdots, \tag{46}$$

$$J^M = \lambda_{\max}^M \left|\lambda_{\max}^M\right\rangle\left\langle\lambda_{\max}^M\right| + \lambda_{2\mathrm{nd}}^M \left|\lambda_{2\mathrm{nd}}^M\right\rangle\left\langle\lambda_{2\mathrm{nd}}^M\right| + \lambda_{3\mathrm{rd}}^M \left|\lambda_{3\mathrm{rd}}^M\right\rangle\left\langle\lambda_{3\mathrm{rd}}^M\right| + \cdots. \tag{47}$$

We show the direction cosines for the four eigenvectors in table 2. From this table, we obtain the following approximate expressions for pattern vectors by eigenvectors:

$$\boldsymbol{\xi}^1 \simeq 0.97\left|\lambda_{\max}^H\right\rangle, \qquad \boldsymbol{\xi}^2 \simeq 0.44\left|\lambda_{2\mathrm{nd}}^H\right\rangle + 0.7\left|\lambda_{3\mathrm{rd}}^H\right\rangle,$$

$$\boldsymbol{\xi}^3 \simeq -0.72\left|\lambda_{2\mathrm{nd}}^H\right\rangle + 0.43\left|\lambda_{3\mathrm{rd}}^H\right\rangle, \qquad \boldsymbol{\xi}^{\mathrm{mix}} \simeq 0.5\left|\lambda_{\max}^H\right\rangle + 0.53\left|\lambda_{3\mathrm{rd}}^H\right\rangle,$$

$$\boldsymbol{\xi}^1 \simeq -0.56\left|\lambda_{\max}^M\right\rangle + 0.65\left|\lambda_{3\mathrm{rd}}^M\right\rangle, \qquad \boldsymbol{\xi}^2 \simeq -0.47\left|\lambda_{\max}^M\right\rangle + 0.55\left|\lambda_{2\mathrm{nd}}^M\right\rangle - 0.48\left|\lambda_{3\mathrm{rd}}^M\right\rangle$$

$$\boldsymbol{\xi}^3 \simeq -0.56\left|\lambda_{\max}^M\right\rangle - 0.61\left|\lambda_{2\mathrm{nd}}^M\right\rangle - 0.25\left|\lambda_{3\mathrm{rd}}^M\right\rangle, \qquad \boldsymbol{\xi}^{\mathrm{mix}} \simeq -0.95\left|\lambda_{\max}^M\right\rangle.$$

Thus, $\left\{J_{ij}^H\right\}$ and $\left\{J_{ij}^M\right\}$ are characterized by three eigenvectors belonging to three large eigenvalues. Contrary to our expectation, direction cosines between eigenvectors of $J^H$ and

**Table 2.** $\tilde{T} = 0.1, \varepsilon = 1.0, T = 0.4, N = 500$. The direction cosines and angles between the eigenvectors and pattern vectors, and eigenvalues. The left and right values in parentheses are direction cosines and angles, respectively.

| | $\boldsymbol{\xi}^1$ | $\boldsymbol{\xi}^2$ | $\boldsymbol{\xi}^3$ | $\boldsymbol{\xi}^{\mathrm{mix}}$ | $\lambda$ |
|---|---|---|---|---|---|
| $\lvert\lambda^H_{\mathrm{max}}\rangle$ | (0.97, 12) | (−0.04, 88) | (0.03, 88) | (0.5, 60) | 1.6 |
| $\lvert\lambda^H_{\mathrm{2nd}}\rangle$ | (0.004, 90) | (0.44, 64) | (−0.72, 44) | (−0.17, 80) | 0.77 |
| $\lvert\lambda^H_{\mathrm{3rd}}\rangle$ | (−0.007, 90) | (0.70, 45) | (0.43, 65) | (0.53, 58) | 0.73 |
| $\lvert\lambda^H_{\mathrm{4th}}\rangle$ | (−0.01, 89) | (−0.12, 83) | (−0.05, 87) | (−0.09, 85) | 0.63 |
| $\lvert\lambda^M_{\mathrm{max}}\rangle$ | (−0.56, 56) | (−0.47, 62) | (−0.56, 56) | (−0.95, 18) | 1.5 |
| $\lvert\lambda^M_{\mathrm{2nd}}\rangle$ | (0.12, 83) | (0.55, 56) | (−0.61, 53) | (0.006, 90) | 0.77 |
| $\lvert\lambda^M_{\mathrm{3rd}}\rangle$ | (0.65, 49) | (−0.48, 61) | (−0.25, 76) | (−0.02, 89) | 0.73 |
| $\lvert\lambda^M_{\mathrm{4th}}\rangle$ | (0.008, 90) | (−0.04, 88) | (0.03, 88) | (−0.03, 88) | 0.63 |

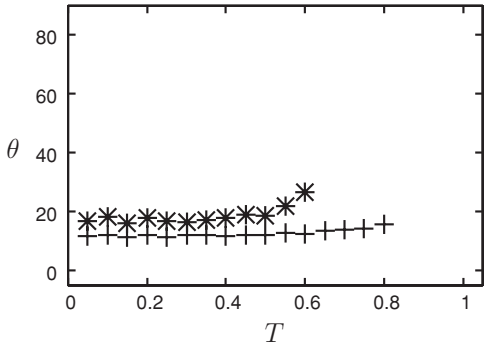


**Figure 10.** The temperature dependence of angles between eigenvector and pattern vector. $\tilde{T} = 0.1, \varepsilon = 1.0, N = 500$. $+$ : $|\lambda^H_{\mathrm{max}}\rangle$ and $\xi^1$, $*$ : $|\lambda^M_{\mathrm{max}}\rangle$ and $\xi^{\mathrm{mix}}$.

$\boldsymbol{\xi}^2$, $\boldsymbol{\xi}^3$ and $\boldsymbol{\xi}^{\mathrm{mix}}$ are not very large. Also, direction cosines between eigenvectors of $J^M$ and $\boldsymbol{\xi}^1$, $\boldsymbol{\xi}^2$ and $\boldsymbol{\xi}^3$ are not very large. This is the reason that there is no coexistence of attractors in $J^H$ and $J^M$. In this respect, our simulations were not sufficient, although the agreement between theoretical and numerical results for $m_\mu$ is excellent as shown in figures 4, 5 and 7. It seems very difficult to obtain interactions by simulations, in which the coexistence of attractors occurs for such a large system size $N$ that theoretical and numerical results agree fairly well.

Now, let us study the temperature dependence of the angle between $\lvert\lambda^H_{\mathrm{max}}\rangle$ and $\boldsymbol{\xi}^1$, and that between $\lvert\lambda^M_{\mathrm{max}}\rangle$ and $\boldsymbol{\xi}^{\mathrm{mix}}$. See figure 10. As seen from the figure, the angles are rather small and about 12° or 18° for $\{J^H_{ij}\}$ or $\{J^M_{ij}\}$ at all temperatures where the Hopfield or the mixed state appears, respectively.

From the results obtained in this section, we conclude that each interaction obtained in our simulations is characterized mainly by the eigenvector $\lvert\lambda^H_{\mathrm{max}}\rangle$ or $\lvert\lambda^M_{\mathrm{max}}\rangle$ for $\{J^H_{ij}\}$ or $\{J^M_{ij}\}$, respectively.

## 5. Summary and discussion

We investigated the change of the system behavior by partial annealing in which the synaptic weights change but much slower than the neurons. We assumed that the system has already

learned $p$ patterns, the Hopfield model, and introduced the coefficient $\varepsilon$ of the Hebbian learning. We studied the stationary states of the Langevin equation by changing parameters, those are the learning coefficient $\varepsilon$, the neuron 'temperature' $T$ and the synaptic weight 'temperature' $\tilde{T}$.

We drew the phase diagrams in the $(T, \varepsilon)$ plane for $\tilde{T} = 0.1$ and 0.4 by taking into account the AT stability. When parameters are changed, various kinds of phase transitions take place. Here, we summarize and discuss significant results among others.

(i) When temperature is decreased from high temperature, the phase transition which takes place firstly is the second-order one for small values of $\varepsilon$ and is the first-order one for large values of $\varepsilon$.
(ii) When temperature is decreased from high temperature, the Hopfield attractor appears from the paramagnetic phase by the first-order phase transition at $T = T_H^{1\mathrm{st}}$ for large values of $\varepsilon$.
(iii) As $\varepsilon$ is increased, the first-order phase transition temperature from the paramagnetic phase to the Hopfield attractor, $T_H^{1\mathrm{st}}$, increases and then the temperature region where the Hopfield attractor is stable increases.

From these results, we can conclude that for large values of $\varepsilon$, the larger the Hebbian learning term $\varepsilon\langle\sigma_i\sigma_j\rangle_{sp}$ is, the more stable the Hopfield attractor is.

When $\tilde{T} = 0.1$, we found that there is the wide region where the Hopfield attractor and the mixed state coexist in the $(T, \varepsilon)$ plane, and the ratio of the temperature region where the mixed state exists to that where the Hopfield attractor exists increases as $\varepsilon$ increases. That is, the learning term not only makes the Hopfield attractor stable but also the mixed state more stable when the synaptic noise $\tilde{T}$ is small and $\varepsilon$ is large.

For large values of $\tilde{T} = 0.4$, we found that there is a temperature region where the spin glass phase exists for small values of $\varepsilon$, whereas for large $\varepsilon$, the spin glass phase disappears, and instead, the Hopfield attractor appears. In addition, the replica symmetry breaking of the Hopfield attractor takes place when temperature is low.

If $\varepsilon = 0$, the present model becomes the Sherrington–Kirkpatrick model (SK model) of spin glass in which the average of the interaction depends on the suffix $ij$ and the variance of $J_{ij}$ is $\frac{\tilde{T}}{N\mu}$. Therefore, it is quite natural that when $\tilde{T}$ is large and the effect of learning is small for small $\varepsilon$, the system tends to be spin glass state. In the SK model, the replica symmetry breaks infinitely when $T$ is small. So, it is reasonable that the replica symmetric Hopfield attractor becomes AT unstable at low value of $T$.

We performed the direct simulations of the Langevin equation and calculated order parameters. The theoretical results were confirmed by the numerical simulations, except for the finite size effects observed only for several values of parameters.

In order to study the nature of interactions generated by partial annealing, we performed the Monte Carlo simulations at $\tilde{T} = 0.1$ and $\varepsilon = 1.0$ using the synaptic weights obtained by partial annealing and found that there are two types of interactions, one of which is $\{J_{ij}^H\}$ in which the Hopfield attractor appears during partial annealing, and the other of which is $\{J_{ij}^M\}$ in which the mixed state appears. Using these two interactions, we performed Monte Carlo simulations with various initial conditions, and found that irrespective of the initial conditions, the neuron system converges to the attractor which appeared in the process of partial annealing. We studied the nature of $\{J_{ij}^H\}$ and $\{J_{ij}^M\}$ in details. We calculated the direction cosines and angles between two interactions, their eigenvalue spectra, eigenvectors, and the direction cosines and angles between eigenvectors and pattern vectors. From these calculations, we found that $\boldsymbol{J}^H$ and $\boldsymbol{J}^M$ have almost the same direction and they are almost perpendicular to $\boldsymbol{K}$, and their eigenvalues are almost the same except for the largest ones. By

the investigation of eigenvectors, we found that the eigenvectors belonging to the three largest eigenvalues have large direction cosines with some of pattern vectors, whereas the direction cosines for the other eigenvectors with pattern vectors are almost 0. Thus, $\left\{J_{ij}^{H}\right\}$ and $\{J_{ij}^{M}\}$ are characterized by three eigenvectors belonging to three large eigenvalues. Further, we studied the temperature dependence of the angle between $\left|\lambda_{\max}^{H}\right\rangle$ and $\boldsymbol{\xi}^{1}$, and that between $\left|\lambda_{\max}^{M}\right\rangle$ and $\boldsymbol{\xi}^{\mathrm{mix}}$, and found that these angles are small at all temperatures where the Hopfield and the mixed state appear, respectively. Therefore, we conclude that each interaction obtained in our simulations is characterized mainly by the eigenvector $\left|\lambda_{\max}^{H}\right\rangle$ or $\left|\lambda_{\max}^{M}\right\rangle$ for $\left\{J_{ij}^{H}\right\}$ or $\left\{J_{ij}^{M}\right\}$, respectively.

Contrary to our expectation, direction cosines between eigenvectors of $J^{H}$ and the pattern vectors $\boldsymbol{\xi}^{2}$, $\boldsymbol{\xi}^{3}$ and $\boldsymbol{\xi}^{\mathrm{mix}}$ are not very large. Neither are direction cosines between eigenvectors of $J^{M}$ and $\boldsymbol{\xi}^{1}$, $\boldsymbol{\xi}^{2}$ and $\boldsymbol{\xi}^{3}$. This is the reason why there is no coexistence of attractors in $J^{H}$ and $J^{M}$. Therefore, our simulations were not sufficient in the sense that the resultant interaction does not have any coexistent attractors, although the theoretical predictions and numerical results on the temperature dependences of order parameters and on the phase transition temperatures agree quite well. Here, let us consider why the theoretical results and numerical ones for the temperature dependence of $m_1$, $m_2$ and $m_3$ agree although the coexistence of attractors does no take place.

Since the learning rule is the Hebbian rule, once an attractor, say $\left\{\sigma_{i}^{A}\right\}$, appears in the process of partial annealing, the relative weight of $\sigma_{i}^{A}\sigma_{j}^{A}$ in the interaction $J_{ij}$ is increased and the stability of the attractor increases. Since the external noise term exists, the learning term is deteriorated, and depending on the values of $T$, $\varepsilon$ and $\tilde{T}$, the degree of stability of the attractor such as the width and depth of its basin is determined. If one could perform a simulation of the Langevin equation, in which a trajectory of $\{J_{ij}\}$ would wander around any region of the phase space of interactions, any attractor would appear one after the other. Let us call such a simulation as the ergodic simulation. As is expected by our simulation results, the resultant interaction obtained by the ergodic simulation would be the summation over such interactions each of which is characterized by one particular pattern vector. Since the attractors are only a few, the interference between attractors would not take place, and then when one of attractors, say $\boldsymbol{\sigma}^{B}$, is retrieved, the temperature dependence of $m_1$, $m_2$ and $m_3$ for the resultant interaction by the ergodic simulation would be the same as that of the interaction obtained by our simulation which has the attractor $\boldsymbol{\sigma}^{B}$.

It would be necessary to perform numerical simulations in an appropriate method to realize the coexistence of attractors. However, it would be very difficult, because we have to attain ergodic behavior in the phase space of the interaction $\{J_{ij}\}$, and at the same time the system size should be large enough in order that the standard deviations for numerical results are fairly small. This is a difficult feature problem.

In the present model, the case of $\tilde{T} = 0$ and $\varepsilon = 0$ is nothing but the original Hopfield model. In this case, the Hopfield attractor and the mixed state are stable for low temperatures. We numerically confirmed that the Hopfield attractor is stable for all values of $\varepsilon$ when it exists for $\tilde{T} = 0.1$, although we cannot ignore a possibility that the AT instability would take place at very low temperature $T$ for $\tilde{T} > 0$. On the other hand, for $\tilde{T} = 0.4$ we found the region in which the Hopfield attractor becomes AT unstable with the replicon mode eigenvalue $\lambda_3 > 0$ indicating replica symmetry breaking. This happens when $\varepsilon$ is less than 1. Thus, it is concluded that if $\varepsilon$ is larger than some value which depends on $\tilde{T}$, the larger $\varepsilon$ is, the wider the temperature region of the stable Hopfield attractor is.

From figure 1, we note that for small values of $\varepsilon$, partial annealing cannot widen the stable region of the Hopfield attractor. In particular, it seems that when $\varepsilon$ becomes negative, i.e. in the case of unlearning, no significant change happens in the phase diagram. We consider that

this is because the loading rate of patterns, $\alpha = \frac{p}{N}$, is 0 in this study. If $\alpha$ is positive and large, results might change. This is an interesting unsolved problem.

**Acknowledgments**

The authors are grateful to Dr Kimoto and Dr Kawamura for valuable discussions. This work is supported by the Nara Women's University Intramural Grant for project research.

## Appendix. AT stability of solutions

In this appendix, the eigenvalues and eigenvectors of the Hessian are calculated. Below, the symbols $\alpha$, $\beta$, $\gamma$ and $\delta$ indicate the replica indexes and symbols $\mu$ and $\nu$ the pattern induces.

Let the Hessian matrix $\mathcal{G}$ be an $L \times L$ matrix. $L$ is the summation of the dimensions of $\left\{\epsilon_\nu^\alpha\right\}$ space and that of $\{\eta^{\alpha\beta}\}$ space, i.e. $L = 3n + {}_nC_2 = \frac{n(n+5)}{2}$ for $p = 3$. The eigenvalue equation is expressed as

$$\mathcal{G}\boldsymbol{\mu} = \lambda\boldsymbol{\mu}, \tag{A.1}$$

where $\lambda$ is an eigenvalue of $\mathcal{G}$ and $\boldsymbol{\mu}$ is the eigenvector belonging to $\lambda$. $\boldsymbol{\mu}$ takes the following form:

$$\boldsymbol{\mu} = \begin{pmatrix} \left\{\epsilon_\nu^\alpha\right\} \\ \{\eta^{\alpha\beta}\} \end{pmatrix}. \tag{A.2}$$

$\left\{\epsilon_\nu^\alpha\right\}$ and $\{\eta^{\alpha\beta}\}$ denote a $3n$ dimensional and a ${}_nC_2$ dimensional column vectors, respectively. Below, $[\cdots]$ is the average over patterns $\xi$ and a bar denotes the following average:

$$\overline{f(\Xi)} \equiv \Omega^{-1} \int Dx \cosh^n(\Xi) f(\Xi), \tag{A.3}$$

$$\Omega \equiv \int Dx \cosh^n(\Xi), \tag{A.4}$$

$$\Xi = \beta\left\{\sqrt{\frac{q}{\mu\tilde{\beta}}}x + \frac{K}{\mu\sqrt{p}}\sum_\nu m_\nu \xi^\nu\right\} = \sqrt{\kappa q}x + \beta J \sum_\nu m_\nu \xi^\nu, \tag{A.5}$$

$$\kappa \equiv \frac{\beta^2}{\mu\tilde{\beta}}, \qquad J \equiv \frac{K}{\mu\sqrt{p}}. \tag{A.6}$$

Further, $\langle\cdots\rangle$ denotes the following average at the replica symmetric solution:

$$\begin{aligned} \langle\cdots\rangle &= \frac{\mathrm{Tr}_{\boldsymbol{\sigma}} \mathrm{e}^{\tilde{H}} \cdots}{\mathrm{Tr}_{\boldsymbol{\sigma}} \mathrm{e}^{\tilde{H}}}, \\ \tilde{H} &= \kappa \sum_{\alpha<\beta} (q_{\alpha\beta})^2 \sigma^\alpha \sigma^\beta + \beta J \sum_{\alpha\nu} m_\nu^\alpha \sigma^\alpha \xi^\nu \\ &= \kappa q^2 \sum_{\alpha<\beta} \sigma^\alpha \sigma^\beta + \beta J \sum_{\alpha\nu} m_\nu \sigma^\alpha \xi^\nu. \end{aligned}$$

### *A.1. Hopfield attractor*

We put $m_1 = m, m_2 = m_3 = 0$. Then the non-zero elements of $\mathcal{G}$ are the seven quantities defined as

$$\mathcal{G}_{(\alpha\nu)(\alpha\nu)} \equiv \frac{\partial^2 G}{\partial m_\nu^{\alpha 2}} = -\beta J(1-\beta J(1-m^2)) = A, \tag{A.7}$$

$$\mathcal{G}_{(\alpha\nu)(\beta\nu)} \equiv \frac{\partial^2 G}{\partial m_\nu^\alpha \partial m_\nu^\beta} = (\beta J)^2(q-m^2) = B, \qquad (\alpha \neq \beta), \tag{A.8}$$

$$\begin{aligned}\mathcal{G}_{(\alpha\beta)(\alpha\beta)} &\equiv \frac{\partial^2 G}{\partial q_{\alpha\beta}\partial q_{\alpha\beta}} = -\kappa[1-\kappa(1-\langle\sigma^\alpha\sigma^\beta\rangle^2)] \\ &= -\kappa[1-\kappa(1-q^2)] = P, \qquad (\alpha \neq \beta).\end{aligned} \tag{A.9}$$

$$\begin{aligned}\mathcal{G}_{(\alpha\beta)(\alpha\gamma)} &\equiv \frac{\partial^2 G}{\partial q_{\alpha\beta}\partial q_{\alpha\gamma}} = \kappa^2[\langle\sigma^\beta\sigma^\gamma\rangle - \langle\sigma^\alpha\sigma^\beta\rangle^2] \\ &= \kappa^2(q-q^2) = Q, \alpha, \beta \text{ and } \gamma \text{are all different.}\end{aligned} \tag{A.10}$$

$$\mathcal{G}_{(\alpha\beta)(\gamma\delta)} \equiv \frac{\partial^2 G}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}} = \kappa^2[\langle\sigma^\alpha\sigma^\beta\sigma^\gamma\sigma^\delta\rangle - \langle\sigma^\alpha\sigma^\beta\rangle^2] = \kappa^2(\overline{\tanh^4\Xi_1} - q^2) = R, \tag{A.11}$$

$\alpha, \beta, \gamma$ and $\delta$ are all different.

$$\mathcal{G}_{(\alpha\beta)(\alpha\nu)} \equiv \frac{\partial^2 G}{\partial q_{\alpha\beta}\partial m_\nu^\alpha} = \kappa\beta Jm(1-q)\delta_{\nu 1} = C\delta_{\nu 1}, \qquad (\alpha \neq \beta). \tag{A.12}$$

$$\mathcal{G}_{(\beta\gamma)(\alpha\nu)} \equiv \frac{\partial^2 G}{\partial q_{\beta\gamma}\partial m_\nu^\alpha} = \kappa\beta J(\overline{\tanh^3\Xi_1} - qm)\delta_{\nu 1} = D\delta_{\nu 1}, \qquad (\alpha \neq \beta, \alpha \neq \gamma). \tag{A.13}$$

Here, $\Xi_1 = \sqrt{\kappa q}x + \beta Jm$. We list the eigenvalues, their degeneracies and eigenvectors:

$$\begin{aligned}&\lambda_1^{(1)\pm} = \frac{1}{2}\{X \pm \sqrt{Y^2+Z}\}, \quad \text{degeneracy: 1 for each,} \\ &\qquad X = A + (n-1)B + P + 2(n-2)Q + \frac{(n-2)(n-3)}{2}R, \\ &\qquad Y = A + (n-1)B - P - 2(n-2)Q - \frac{(n-2)(n-3)}{2}R, \\ &\qquad Z = 2(n-1)\{2C + (n-2)D\}^2, \\ &\qquad \epsilon_1^\alpha = a, \quad \epsilon_2^\alpha = 0, \quad \epsilon_3^\alpha = 0, \quad \eta^{\alpha\beta} = b;\end{aligned} \tag{A.14}$$

$$\begin{aligned}&\lambda_1^{(2)} = A + (n-1)B, \quad \text{degeneracy: 2,} \\ &\qquad \epsilon_1^\alpha = 0, \quad \epsilon_\mu^\alpha = a', \quad \eta^{\alpha\beta} = 0, \quad \mu = 2 \text{ or } 3;\end{aligned} \tag{A.15}$$

$$\begin{aligned}&\lambda_2^{(1)\pm} = \frac{1}{2}\{X' \pm \sqrt{(Y')^2+Z'}\}, \quad \text{degeneracy : } (n-1) \text{ for each,} \\ &\qquad X' = A - B + P + (n-4)Q - (n-3)R, \\ &\qquad Y' = A - B - P - (n-4)Q + (n-3)R, \\ &\qquad Z' = 4(n-2)(C-D)^2, \\ &\qquad \epsilon_1^\theta = c_1, \quad \epsilon_1^\alpha = d_1, \quad \epsilon_2^\beta = 0, \quad \epsilon_3^\beta = 0, \quad \alpha \neq \theta, \quad \theta \text{ is some replica index,} \\ &\qquad \eta^{\theta\beta} = \eta^{\alpha\theta} = f, \quad \eta^{\alpha\beta} = g, \quad \alpha \neq \theta \quad \text{and} \quad \beta \neq \theta;\end{aligned} \tag{A.16}$$

$$\begin{aligned}&\lambda_2^{(2)} = A - B, \quad \text{degeneracy : } 2(n-1), \\ &\qquad \epsilon_\mu^\theta = c_2, \quad \epsilon_\mu^\alpha = d_2, \quad \epsilon_\nu^\beta = 0, \quad \alpha \neq \theta, \quad \nu \neq \mu, \\ &\qquad \theta \text{ is some replica index,} \quad \mu = 2 \text{ or } 3, \\ &\qquad \eta^{\alpha\beta} = 0, \quad \text{for any } \alpha, \beta;\end{aligned} \tag{A.17}$$

$$\begin{aligned}
&\lambda_3 = P - 2Q + R, \quad \text{degeneracy} : \frac{n(n-1)}{2} - n, \\
&\quad \epsilon_1^{\theta_1} = \epsilon_1^{\theta_2} = h_1, \quad \epsilon_1^{\alpha} = h_2, \quad \epsilon_\nu^{\beta} = 0, \quad \theta_1 \neq \theta_2, \\
&\quad \alpha \neq \theta_1, \quad \alpha \neq \theta_2, \quad \nu \neq 1, \\
&\quad \theta_1 \text{ and } \theta_2 \text{ are some replica indexes,}
\end{aligned} \tag{A.18}$$

$$\begin{aligned}
&\eta^{\theta_1\theta_2} = u, \quad \eta^{\theta_1\alpha} = \eta^{\theta_2\alpha} = v, \quad \eta^{\alpha\beta} = w, \\
&\quad \alpha \neq \theta_1, \quad \alpha \neq \theta_2 \quad \text{and} \quad \beta \neq \theta_1, \quad \beta \neq \theta_2.
\end{aligned} \tag{A.19}$$

*A.2. Mixed state with three patterns*

We put $m_1 = m_2 = m_3 = m$. $\Xi_m$ is defined by

$$\Xi = \sqrt{\kappa q}x + \beta J \sum_\nu m_\nu \xi_\nu = \sqrt{\kappa q}x + \beta J m(\xi_1 + \xi_2 + \xi_3) \equiv \Xi_m. \tag{A.20}$$

Now, we list the non-zero elements of $\mathcal{G}$:

$$\mathcal{G}_{(\alpha\nu)(\alpha\nu)} \equiv \frac{\partial^2 G}{\partial m_\nu^{\alpha 2}} = -\beta J + (\beta J)^2\{1 - [(\overline{\tanh \Xi_m})^2]\} = A_1. \tag{A.21}$$

Since this quantity does not depend on $\nu$, we define it $A_1$. Below, we assume $\mu \neq \nu$ and $\alpha, \beta, \gamma$ and $\delta$ are all different:

$$\begin{aligned}
\mathcal{G}_{(\alpha\nu)(\alpha\mu)} &\equiv \frac{\partial^2 G}{\partial m_\nu^\alpha \partial m_\mu^\alpha} = (\beta J)^2\{[\langle \sigma^\alpha \sigma^\beta \xi^\nu \xi^\mu \rangle] - [\langle \sigma^\alpha \xi^\nu \rangle \langle \sigma^\alpha \xi^\mu \rangle]\} \\
&= (\beta J)^2\{[\xi^\nu \xi^\mu] - [\langle \sigma^\alpha \rangle^2 \xi^\mu \xi^\nu]\} \\
&= -(\beta J)^2[(\overline{\tanh \Xi_m})^2 \xi^\mu \xi^\nu] = A_2,
\end{aligned} \tag{A.22}$$

$$\begin{aligned}
\mathcal{G}_{(\alpha\nu)(\beta\nu)} &\equiv \frac{\partial^2 G}{\partial m_\nu^\alpha \partial m_\nu^\beta} = (\beta J)^2\{[\langle \sigma^\alpha \sigma^\beta \rangle] - [\langle \sigma^\alpha \rangle \xi^\nu \langle \sigma^\beta \rangle \xi^\nu]\} \\
&= (\beta J)^2\{[\overline{\tanh^2 \Xi_m}] - [(\overline{\tanh \Xi_m})^2]\} \\
&= (\beta J)^2\{q - [(\overline{\tanh \Xi_m})^2]\} = B_1.
\end{aligned} \tag{A.23}$$

At the last line, we used the following relations:

$$q = [\langle \sigma^\alpha \sigma^\beta \rangle] = [\overline{\tanh^2 \Xi_m}], \tag{A.24}$$

$$m_\nu^\alpha = [\langle \sigma^\alpha \rangle \xi^\nu] = [\overline{\tanh \Xi_m} \xi^\nu] = m_\nu = m, \tag{A.25}$$

$$\begin{aligned}
\mathcal{G}_{(\alpha\nu)(\beta\mu)} &\equiv \frac{\partial^2 G}{\partial m_\nu^\alpha \partial m_\mu^\beta} = (\beta J)^2\{[\langle \sigma^\alpha \sigma^\beta \rangle \xi^\nu \xi^\mu] - [\langle \sigma^\alpha \rangle \langle \sigma^\beta \rangle \xi^\nu \xi^\mu]\} \\
&= (\beta J)^2\{\overline{\tanh^2 \Xi_m} \xi^\nu \xi^\mu - [(\overline{\tanh \Xi_m})^2 \xi^\mu \xi^\nu]\} = B_2,
\end{aligned} \tag{A.26}$$

$$\begin{aligned}
\mathcal{G}_{(\alpha\beta)(\alpha\nu)} &\equiv \frac{\partial^2 G}{\partial q_{\alpha\beta} \partial m_\nu^\alpha} = \kappa\beta J[\langle \sigma^\alpha \sigma^\beta \sigma^\alpha \xi^\nu \rangle - \langle \sigma^\alpha \sigma^\beta \rangle \langle \sigma^\alpha \xi^\nu \rangle] \\
&= \kappa\beta J\{[\overline{\tanh \Xi_m} \xi^\nu] - [\overline{\tanh^2 \Xi_m} \cdot \overline{\tanh \Xi_m} \xi^\nu]\} \\
&= \kappa\beta J\{m - [\overline{\tanh^2 \Xi_m} \cdot \overline{\tanh \Xi_m} \xi^\nu]\} \equiv C,
\end{aligned} \tag{A.27}$$

$$\mathcal{G}_{(\alpha\beta)(\nu\gamma)} \equiv \frac{\partial^2 G}{\partial q_{\alpha\beta}\partial m_\nu^\gamma} = \kappa\beta J[\langle\sigma^\alpha\sigma^\beta\sigma^\gamma\rangle\xi^\nu - \langle\sigma^\alpha\sigma^\beta\rangle\langle\sigma^\gamma\rangle\xi^\nu]$$
$$= \kappa\beta J\{\overline{[\tanh^3\Xi_m}\xi^\nu] - \overline{[\tanh^2\Xi_m}\cdot\overline{\tanh\Xi_m}\xi^\nu]\} \equiv D, \tag{A.28}$$

$$\mathcal{G}_{(\alpha\beta)(\alpha\beta)} \equiv \frac{\partial^2 G}{\partial q_{\alpha\beta}^2} = -\kappa + \kappa^2\{1 - [\langle\sigma^\alpha\sigma^\beta\rangle^2]\}$$
$$= -\kappa + \kappa^2\{1 - \overline{[(\tanh^2\Xi_m)^2]}\} \equiv P, \tag{A.29}$$

$$\mathcal{G}_{(\alpha\beta)(\alpha\gamma)} \equiv \frac{\partial^2 G}{\partial q_{\alpha\beta}\partial q_{\alpha\gamma}} = \kappa^2[\langle\sigma^\alpha\sigma^\beta\sigma^\alpha\sigma^\gamma\rangle - \langle\sigma^\alpha\sigma^\beta\rangle^2]$$
$$= \kappa^2\{\overline{[\tanh^2\Xi_m]} - \overline{[(\tanh^2\Xi_m)^2]}\}$$
$$= \kappa^2\{q - \overline{[(\tanh^2\Xi_m)^2]}\} \equiv Q, \tag{A.30}$$

$$\mathcal{G}_{(\alpha\beta)(\gamma\delta)} \equiv \frac{\partial^2 G}{\partial q_{\alpha\beta}\partial q_{\gamma\delta}} = \kappa^2[\langle\sigma^\alpha\sigma^\beta\sigma^\gamma\sigma^\delta\rangle - \langle\sigma^\alpha\sigma^\beta\rangle^2]$$
$$= \kappa^2\{\overline{[\tanh^4\Xi_m]} - \overline{[(\tanh^2\Xi_m)^2]}\} \equiv R. \tag{A.31}$$

We list the eigenvalues, their degeneracies and eigenvectors:

$$\lambda_1^{(1)\pm} = \frac{1}{2}\{X \pm \sqrt{Y^2 + Z}\}, \qquad \text{degeneracy: 1 for each,}$$
$$X = A + (n-1)B + P + 2(n-2)Q + \frac{(n-2)(n-3)}{2}R,$$
$$Y = A + (n-1)B - P - 2(n-2)Q - \frac{(n-2)(n-3)}{2}R,$$
$$Z = 6(n-1)\big\{2C + (n-2)D\big\}^2,$$
$$A = A_1 + (n-1)B_1, \qquad B = \frac{2(A_2 + (n-1)B_2)}{n-1},$$
$$\epsilon_1^\alpha = \epsilon_2^\alpha = \epsilon_3^\alpha = a, \qquad \eta^{\alpha\beta} = b, \tag{A.32}$$

$$\lambda_1^{(2)} = A_1 - A_2 + (n-1)(B_1 - B_2), \quad \text{degeneracy: 2,}$$
$$\epsilon_1^\alpha = a_1', \quad \epsilon_2^\alpha = a_2', \quad \epsilon_3^\alpha = a_3', \quad a_1' + a_2' + a_3' = 0, \quad \eta^{\alpha\beta} = 0, \tag{A.33}$$

$$\lambda_2^{(1)\pm} = \frac{1}{2}\{X' \pm \sqrt{(Y')^2 + Z'}\}, \quad \text{degeneracy : } (n-1) \text{ for each,}$$
$$X' = A_1 - B_1 + 2(A_2 - B_2) + P + (n-4)Q - (n-3)R,$$
$$Y' = A_1 - B_1 + 2(A_2 - B_2) - P - (n-4)Q + (n-3)R,$$
$$Z' = 12(n-2)(C-D)^2,$$
$$\epsilon_1^\theta = c_1, \quad \epsilon_1^\alpha = d_1, \quad \epsilon_2^\theta = c_2, \quad \epsilon_2^\alpha = d_2, \quad \epsilon_3^\theta = c_3, \quad \epsilon_3^\alpha = d_3,$$
$$\alpha \neq \theta, \theta \text{ is some replica index,}$$
$$\eta^{\theta\beta} = \eta^{\alpha\theta} =, \quad \eta^{\alpha\beta} = g, \quad \alpha \neq \theta \quad \text{and} \quad \beta \neq \theta, \tag{A.34}$$

$$\begin{aligned}
&\lambda_2^{(2)} = A_1 - A_2 - (B_1 - B_2), \quad \text{degeneracy} : 2(n-1), \quad \epsilon_1^{\theta} = c_1, \\
&\quad \epsilon_1^{\alpha} = d_1, \quad \epsilon_2^{\theta} = c_2, \quad \epsilon_2^{\alpha} = d_2, \quad \epsilon_3^{\theta} = c_3, \quad \epsilon_3^{\alpha} = d_3, \\
&\quad \alpha \neq \theta, \quad \theta \text{ is some replica index}, \\
&\quad \eta^{\alpha\beta} = 0, \quad \text{for any } \alpha, \beta, \\
&\quad \eta^{\alpha\beta} = 0, \quad \text{for any } \alpha, \beta, \\
&\quad c_i = (1-n)d_i, \quad d_1 + d_2 + d_3 = 0,
\end{aligned} \tag{A.35}$$

$$\begin{aligned}
&\lambda_3 = P - 2Q + R, \quad \text{degeneracy} : \frac{n(n-1)}{2} - n, \\
&\quad \epsilon_{\mu}^{\theta_1} = \epsilon_{\mu}^{\theta_2} = r_{\mu}, \quad \epsilon_{\mu}^{\alpha} = s_{\mu}, \quad \mu = 1, 2, 3, \quad \theta_1 \neq \theta_2, \quad \alpha \neq \theta_1, \quad \alpha \neq \theta_2, \\
&\quad \theta_1 \quad \text{and} \quad \theta_2 \text{ are some replica indexes}, \\
&\quad \eta^{\theta_1\theta_2} = u, \quad \eta^{\theta_1\alpha} = \eta^{\theta_2\alpha} = v, \quad \eta^{\alpha\beta} = w, \\
&\quad \alpha \neq \theta_1, \quad \alpha \neq \theta_2 \quad \text{and} \quad \beta \neq \theta_1, \beta \neq \theta_2.
\end{aligned} \tag{A.36}$$

### *A.3. Spin glass*

For the spin glass solution, $m_{\mu} = 0$ and $q \neq 0$. Thus, we only have to put $m = 0$, $C = D = 0$ in the quantities for the Hopfield attractor. Thus, we obtain the following eigenvalues:

$$\lambda_1^{(1)+} = A + (n-1)B = \lambda_1^{(2)} = \lambda_1^{(3)}, \tag{A.37}$$

$$\lambda_1^{(1)-} = P + 2(n-2)Q + \frac{(n-2)(n-3)}{2}R, \tag{A.38}$$

$$\lambda_2^{(1)+} = A - B, \tag{A.39}$$

$$\lambda_2^{(1)-} = P + (n-1)Q - (n-3)R = \lambda_2^{(2)} = \lambda_2^{(3)}, \tag{A.40}$$

$$\lambda_3 = P - 2Q + R. \tag{A.41}$$

## References


[1] Shinomoto S 1987 *J. Phys. A: Math. Gen.* **20** L1305
[2] Torres J J, Garrido P L and Marro J 1997 *J. Phys. A: Math. Gen.* **30** 7801
[3] Lattanzi G, Nardulli G, Pasquariello G and Stramaglia S 1997 *Phys. Rev.* E **56** 4567
[4] Bressloff P C 1999 *Phys. Rev.* E **60** 2160
[5] Coolen A C C, Penney R W and Sherrington D 1993 *Phys. Rev.* B **48** 16116
[6] Penney R W, Coolen A C C and Sherrington D 1993 *J. Phys. A: Math. Gen.* **26** 3681
[7] Penney R W and Sherrington D 1994 *J. Phys. A: Math. Gen.* **27** 4027
[8] Dotsenko V, Franz S and Mézard M 1994 *J. Phys. A: Math. Gen.* **27** 2351
[9] Uezu T and Coolen A C C 2002 *J. Phys. A: Math. Gen.* **35** 2761
[10] Abe K 2008 *Masters Thesis* Graduate School of Humanities and Sciences, Nara Women's University, Nara (in Japanese)
[11] Hara K, Miyoshi S, Uezu T and Okada M 2008 *Proc. of Int. Conf. on Neural Information Processing*
[12] Kimoto T *et al* in preparation
[13] de Almeida J R L and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983