Unlearning of Mixed States in the Hopfield Model – Finite Loading Case –

Haruka Otani[1], Midori Yoshida[1], Shuji Kiyokawa[2], and Tatsuya Uezu[1] *

[1] *Graduate School of Humanities and Sciences, Nara Women's University, Nara 630, Japan*
[2] *Faculty of Science, Nara Women's University, Nara 630, Japan*

We study the unlearning of mixed states in the Hopfield model for the finite loading case, that is, $\alpha = \frac{p}{N} \ll 1$, where N and p are the numbers of neurons and embedded patterns, respectively. In the general situation that any number of mixed states that exist in the model is unlearned, we derive the saddle point equations (SPEs) and evolution equations for overlaps by introducing sublattices. We postulate a condition that the solutions are stable in equilibrium, and prove that the static and dynamic stabilities are the same. We also prove that the stable state of the Hopfield model continuously changes and is statically and dynamically stable for sufficiently small unlearning coefficients. For $p = 3$, we perform detailed theoretical and numerical calculations. In the case that a single mixed state is unlearned, we determine phase boundaries using the Hessian matrix and by numerically integrating evolution equations. We performed Markov chain Monte Carlo simulations and find that the simulation results agree with the theoretical ones reasonably well. For general p, when all of the mixed states are unlearned with an equal unlearning coefficient η, we derive the formulae for critical unlearning coefficients at the temperature $T = 0$ below which embedded patterns and mixed states exist and are stable as solutions of the SPEs. We found that there is an unlearning region of (T, η) in which all patterns are retained and all mixed states are deleted, although tuning the parameters in this region is more difficult as p increases since the region shrinks. We numerically confirmed the theoretical results of the p dependences of the critical unlearning coefficients.

1. Introduction

One of the roles of dreaming is considered to be to regulate memories, that is, unnecessary memories are considered to be deleted by dreaming.[1)] In this context, we consider strengthening important memories and weakening unnecessary memories. In this work, as a concrete

*uezu@cc.nara-wu.ac.jp

model, we study the Hopfield model.[2] As is well known, in the Hopfield model, when p patterns are stored, a combination of several patterns, a mixed state, is also stored. There are several types of stable mixed state.[3–5] We study the unlearning of mixed states. There have been many studies on unlearning in neural networks.[6–10] These papers treat unlearning the final states starting from noisy inputs, or unlearning by parallel dynamics, or unlearning by using asymmetric synaptic weights, and so forth. In particular, in Refs. 11 and 12, as a candidate of unnecessary memories, a spin glass state is considered.

In this study, we treat symmetric synaptic weights and asynchronous dynamics. In the Hopfield model, spin glass states appear for the extensive loading of patterns. Since we study the finite loading case, mixed states are only spurious states. We investigate the statics and dynamics of the network taking unlearning into account.

We postulate that the stability at the equilibrium is determined by the condition that the para state is stable at high temperatures. We prove that the static stability determined by this assumption agrees with the dynamic stability in general situations. First of all, as the simplest case, we study the case $p = 3$, and perform detailed theoretical and numerical calculations. In particular, for the case that a single mixed state is unlearned, we numerically integrated evolution equations by the Runge-Kutta (RK) method, and performed Markov chain Monte Carlo (MCMC) simulations. The numerical results agree with the theoretical ones quite well. Since we found that the RK method gives the same phase boundaries as those determined by the eigenvalues of the Hessian matrix and MCMC simulations for the unlearning of the single mixed state, we study the unlearning of multiple mixed states by the RK method for general p. Secondly, we study the case that all mixed states are unlearned and derive critical unlearning coefficients at $T = 0$, below which embedded patterns and mixed states exist. We show that we can eliminate all mixed states and retain all patterns if the temperature and the unlearning coefficient take values in some region in (T, η) space. The larger the number of patterns, the smaller the difference between two critical coefficients. In particular, both coefficients are zero at $p = \infty$. This implies that the larger the value of p, the more difficult it is to tune the parameters in the region.

This paper is organized as follows. In sect. 2, we formulate the problem in the general situation that any number of mixed states is unlearned, and describe the saddle point equations (SPEs) of overlaps in the equilibrium and evolution equations for overlaps and sublattice overlaps. In sect. 3, we study in detail the case that one and several mixed states are unlearned for $p = 3$ in detail. In sect. 4, we study the case that all mixed states are unlearned for a general p. Section 5 contains a summary and discussion of the results. In Appendices A and B, we

derive evolution equations for sublattice overlaps and prove the equivalence of the static and dynamic stabilities of the solutions of the SPEs, respectively. In Appendices C and D, in the case that the single mixed state is unlearned, we describe the SPEs and the stability of the solutions of the SPEs, respectively.

2. Formulation

The Hopfield model is a recurrent network of N neurons, in which all neurons interact with each other. The state of the ith neuron is represented by s_i. $s_i = 1$ or $s_i = -1$ corresponds to a firing state or a rest state, respectively. The number of patterns is set to p. Let $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \xi_2^\mu, ..., \xi_N^\mu)$ be μth pattern, where $\mu = 1, 2, \cdots, p$. We assume that ξ_i^μ takes values of ± 1 independently with the probability $1/2$. Let us denote the configuration of N neurons as $\boldsymbol{s} = \{s_i\}$ and let J_{ij} be the synaptic weight from the jth neuron to the ith neuron. We assume $J_{ij} = J_{ji}$ and $J_{ii} = 0$. The synaptic weight of the Hopfield model, $J_{ij}^{(\mathrm{H})}$, is given by

$$\begin{cases} J_{ij}^{(\mathrm{H})} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu, & (i \neq j), \\ J_{ii}^{(\mathrm{H})} = 0. \end{cases} \tag{1}$$

The input signal to the ith neuron at time t, $h_i(t)$, is given by

$$h_i(t) = \sum_{j(\neq i)} J_{ij} s_j(t). \tag{2}$$

In the deterministic update, the new state of the ith neuron is

$$s_i(t + \Delta t) = \mathrm{sgn}\left(\sum_{j(\neq i)} J_{ij} s_j(t)\right), \tag{3}$$

where $\mathrm{sgn}(x) = 1$ for $x \geq 0$ and -1 for $x < 0$. We take the time increment $\Delta t = \frac{1}{N}$. In this study, we consider the probabilistic update, and the probability that the ith neuron takes the value ± 1 is given by

$$\mathrm{Prob}[s_i(t + \Delta t) = \pm 1] = \frac{1 \pm \tanh[\beta h_i(t)]}{2}. \tag{4}$$

Here, $\beta = \frac{1}{T}$ and T represents the strength of noise and is called the 'temperature' in this paper.

2.1 Statics

By adopting an asynchronous update, the stationary state of the network becomes equivalent to the equilibrium state of the canonical ensemble with the following Hamiltonian H:

$$H = -\sum_{i<j} J_{ij} s_i s_j. \tag{5}$$

In the Hopfield model, for $p \ll N$, all patterns are stored when $T < T_c$, where T_c is given by $T_c = 1$. Furthermore, mixed states with three patterns are also stable when $T < T_3$, where $T_3 \simeq 0.46$.[4)]

In this paper, we treat the case of a general p and consider the unlearning of any number of mixed states with three patterns. A mixed state is denoted by $\boldsymbol{\xi}^{(\boldsymbol{\mu};\boldsymbol{\gamma})} = (\xi_1^{(\boldsymbol{\mu};\boldsymbol{\gamma})}, \xi_2^{(\boldsymbol{\mu};\boldsymbol{\gamma})}, \cdots, \xi_N^{(\boldsymbol{\mu};\boldsymbol{\gamma})})$ and determined as

$$\xi_i^{(\boldsymbol{\mu};\boldsymbol{\gamma})} = \mathrm{sgn}(\gamma_1 \xi_i^{\mu_1} + \gamma_2 \xi_i^{\mu_2} + \gamma_3 \xi_i^{\mu_3}),\ i = 1, \cdots, N. \tag{6}$$

Here, $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ and μ_1, μ_2, and μ_3 are all different and take values in $\{1, 2, \cdots, p\}$. $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$ and $|\gamma_i| = 1, (i = 1, 2, 3)$. Without loss of generality, we set $\mu_1 < \mu_2 < \mu_3$. We subtract the mixed states from $J_{ij}^{(\mathrm{H})}$ and treat the following synaptic weight J_{ij},

$$J_{ij} = J_{ij}^{(\mathrm{H})} + J_{ij}^{(\mathrm{U})} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu} + \frac{1}{N} \sum_{\mathcal{V}} \zeta_{(\boldsymbol{\mu};\boldsymbol{\gamma})} \xi_i^{(\boldsymbol{\mu};\boldsymbol{\gamma})} \xi_j^{(\boldsymbol{\mu};\boldsymbol{\gamma})}. \tag{7}$$

Here, $\zeta_{(\boldsymbol{\mu};\boldsymbol{\gamma})}$ is the unlearning coefficient, that is, when $\zeta_{(\boldsymbol{\mu};\boldsymbol{\gamma})}$ is negative, $\xi^{(\boldsymbol{\mu};\boldsymbol{\gamma})}$ is unlearned. $\mathcal{V}$ is the set of superscripts $(\boldsymbol{\mu};\boldsymbol{\gamma})$ with which a mixed state $\xi^{(\boldsymbol{\mu};\boldsymbol{\gamma})}$ is unlearned, and $\sum_{\mathcal{V}}$ denotes the summation of all mixed states with superscripts in the set $\mathcal{V}$. Let u be the number of unlearned mixed states, $u \equiv |\mathcal{V}|$. When all patterns $\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^p$ change their signs, that is, by the transformation $(\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^p) \to (-\boldsymbol{\xi}^1, \cdots, -\boldsymbol{\xi}^p)$, J_{ij} does not change. We call this the inversion symmetry. Because of the inversion symmetry, we may set $\gamma_1 = 1$ without loss of generality. At first sight, Eq. (7) seems to represent the unlearning of the mixed states $\boldsymbol{\xi}^{(\boldsymbol{\mu};\boldsymbol{\gamma})}$ only. However, there is a correlation between the mixed state with $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ and the pattern $\boldsymbol{\xi}^{\mu_k}, (k = 1, 2, 3)$, that is,

$$\frac{1}{N} \sum_{i=1}^{N} \xi_i^{(\boldsymbol{\mu};\boldsymbol{\gamma})} \xi_i^{\mu_k} = \langle\!\langle \xi^{(\boldsymbol{\mu};\boldsymbol{\gamma})} \xi^{\mu_k} \rangle\!\rangle = \frac{1}{2} \gamma_k \text{ for } k = 1, 2, 3. \tag{8}$$

Here, double brackets $\langle\!\langle \cdot \rangle\!\rangle$ denote the average over $\xi^1, \xi^2, \cdots, \xi^p$, where ξ^{μ} takes values of ± 1 with probability 1/2. In this paper, when a quantity is averaged and expressed by double brackets, we omit its subscript, which is i in the present case, if it does not cause any confusion. The first equality follows from the self-averaging because $N \gg 2^p$. Therefore, a part of $\boldsymbol{\xi}^{\mu_k}$ is subtracted from J_{ij} for $k = 1, 2, 3$. Nevertheless, we still call the present procedure the 'unlearning' of the mixed states for simplicity. For later use, we write correlations between two mixed states $\boldsymbol{\xi}^{(\boldsymbol{\mu};\boldsymbol{\gamma})}$ and $\boldsymbol{\xi}^{(\boldsymbol{\mu}';\boldsymbol{\gamma}')}$ where $\boldsymbol{\mu}' = (\mu_1', \mu_2', \mu_3')$ and $\boldsymbol{\gamma}' = (\gamma_1', \gamma_2', \gamma_3')$.

$$\frac{1}{N} \sum_{i} \xi_i^{(\boldsymbol{\mu};\boldsymbol{\gamma})} \xi_i^{(\boldsymbol{\mu}';\boldsymbol{\gamma}')} = \langle\langle \xi^{(\boldsymbol{\mu};\boldsymbol{\gamma})} \xi^{(\boldsymbol{\mu}';\boldsymbol{\gamma}')} \rangle\rangle \tag{9}$$

$$= \begin{cases} 0 & \{\mu_1,\mu_2,\mu_3\}\cap\{\mu_1',\mu_2',\mu_3'\}=\phi, \text{ none of the elements of } \boldsymbol{\mu} \\ & \text{and } \boldsymbol{\mu}' \text{ agree.} \\ \frac{1}{4}\gamma_i\gamma_l', & \mu_i=\mu_l', \text{ one of the elements of } \boldsymbol{\mu} \text{ and } \boldsymbol{\mu}' \text{ agrees.} \\ \frac{1}{4}(\gamma_i\gamma_l'+\gamma_j\gamma_m'), & \mu_i=\mu_l', \mu_j=\mu_m', \text{ two of the elements of } \boldsymbol{\mu} \text{ and } \boldsymbol{\mu}' \text{ agree.} \\ 1 & \boldsymbol{\mu}=\boldsymbol{\mu}', \boldsymbol{\gamma}=\boldsymbol{\gamma}' \\ 0 & \boldsymbol{\mu}=\boldsymbol{\mu}', \boldsymbol{\gamma}\neq\boldsymbol{\gamma}' \end{cases} \tag{10}$$

2.2 SPEs

In this subsection, we formulate the general case that any number of mixed states is unlearned. For simplicity, we denote mixed states by $\boldsymbol{\xi}^{p+1},\cdots,\boldsymbol{\xi}^{v}$, where $v=p+u$, and rewrite J_{ij} [Eq. (7)] as

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{v}\zeta_\mu\xi_i^\mu\xi_j^\mu, \tag{11}$$

where $\zeta_1=\cdots=\zeta_p=1$. The order parameters of the system are the overlaps m^μ between $\boldsymbol{s}$ and $\boldsymbol{\xi}$s,

$$m^\mu(t) = \frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu s_i(t), \qquad (\mu=1,\cdots,v). \tag{12}$$

By the standard recipe, the free energy per neuron f and the SPEs are derived as

$$f = \sum_{\mu=1}^{v}\zeta_\mu\frac{(m^\mu)^2}{2} - \frac{1}{\beta}\langle\!\langle \ln[2\cosh\beta(\sum_{\mu=1}^{v}\zeta_\mu m^\mu\xi^\mu)]\rangle\!\rangle, \tag{13}$$

$$m^\mu = \langle\!\langle \xi^\mu\tanh[\beta(\sum_{\nu=1}^{v}\zeta_\nu m^\nu\xi^\nu)]\rangle\!\rangle, \ \mu=1,\cdots,v. \tag{14}$$

2.3 Stability of solutions of the SPEs

In this subsection, we study the stability of the solutions of the SPEs. The stability of a solution is determined by the Hessian matrix $\Lambda=\{\Lambda_{\mu\nu}\}$, where $\Lambda_{\mu\nu}=\frac{\partial^2 f}{\partial m^\mu\partial m^\nu}$, which is evaluated at the solution. To find a condition under which the solution is stable, we study the behavior of f at high temperatures. For $\beta\ll 1$, f is expressed as

$$f \simeq \sum_{\mu=1}^{v}\zeta_\mu\frac{(m^\mu)^2}{2} + \text{const.} \tag{15}$$

Since the para (P) state in which the m^μ are 0 for all μ should be stable at high temperatures, f should be minimal with respect to m^μ for positive ζ_μ and be maximal with respect to m^μ for negative ζ_μ. Let λ be the eigenvalues of Λ, and let us define λ as $\lambda_1\geq\lambda_2\geq\lambda_3\geq\lambda_4\cdots\geq\lambda_v$. For example, let us consider the case that $p=3$ and only the mixed state $\boldsymbol{\xi}^4\equiv\boldsymbol{\xi}^{(1,2,3;1,1,1)}$ is

unlearned. Defining $\zeta_1 = \zeta_2 = \zeta_3 = 1$ and $\zeta_4 = -\eta$, the condition under which the solution is stable is given by $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0, \lambda_4 < 0$ for $\eta > 0$ and $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0, \lambda_4 > 0$ for $\eta < 0$. Thus, in general situations, a solution is stable when the number of positive (negative) eigenvalues is the same as the number of positive (negative) ζ_μ. We prove that the thus defined static stability agrees with the dynamic stability in Appendix B.

In the next subsection, we formulate the dynamics of the network.

2.4 Dynamics

Let $p_t(\boldsymbol{s})$ be the probability that the configuration of neurons is $\boldsymbol{s} = (s_1, s_2, ..., s_k, ..., s_N)$ at time t and let $w_k(\boldsymbol{s})$ be the transition probability from $\boldsymbol{s}$ to $F_k\boldsymbol{s} \equiv (s_1, s_2, ..., -s_k, ..., s_N)$ per unit of time. F_k is the flip operator of the kth neuron. We adopt the following function as $w_k(\boldsymbol{s})$:

$$w_k(\boldsymbol{s}) = \frac{1}{2}[1 - s_k \tanh(\beta h_k)], \tag{16}$$

$$h_k = \sum_{j(\neq k)} J_{kj} s_j. \tag{17}$$

Then, the time evolution of $p_t(\boldsymbol{s})$ is described by the following master equation:

$$\frac{\partial}{\partial t} p_t(\boldsymbol{s}) = \sum_{k=1}^{N} [w_k(F_k\boldsymbol{s}) p_t(F_k\boldsymbol{s}) - w_k(\boldsymbol{s}) p_t(\boldsymbol{s})]. \tag{18}$$

It is shown that the stationary state of Eq. (18) is the equilibrium state of the canonical ensemble with the Hamiltonian given by Eq. (5).

Now, let us introduce sublattices and sublattice overlaps. The number of possible configurations of $\{\xi_i^\mu\} \equiv (\xi_i^1, \xi_i^2, \cdots, \xi_i^p)$ with i fixed is 2^p. For each configuration, we introduce a sublattice Λ_l of neurons in such a way that $\Lambda_{l+2^{p-1}} = \Lambda_l$ holds for $l = 1, \cdots, 2^{p-1}$. The number of elements in Λ_l, $|\Lambda_l|$, is $|\Lambda_l| = \dfrac{N}{2^p}$ $(l = 1, 2, \cdots, 2^p)$. The sublattice overlap $\mathcal{M}_l$ is defined as

$$\mathcal{M}_l = \frac{1}{|\Lambda_l|} \sum_{i \in \Lambda_l} s_i = \frac{2^p}{N} \sum_{i \in \Lambda_l} s_i, \ l = 1, \cdots, 2^p.$$

Using $\{\mathcal{M}_l\}$, the overlap m^μ is expressed as

$$m^\mu = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu s_i = \frac{1}{N} \sum_{l=1}^{2^p} \sum_{i \in \Lambda_l} \xi_i^\mu s_i = \frac{1}{N} \sum_{l=1}^{2^p} \xi^{\mu,l} \sum_{i \in \Lambda_l} s_i = \frac{1}{2^p} \sum_{l=1}^{2^p} \xi^{\mu,l} \mathcal{M}_l, \ \mu = 1, \cdots, p \tag{19}$$

$\xi^{\mu,l}$ is the value of ξ_i^μ for i in Λ_l. Now, let us study the evolution equations for $\{\mathcal{M}_l\}$. We define $\boldsymbol{\mathcal{M}} = (\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_{2^p})$ and $d\boldsymbol{\mathcal{M}} = \prod_{l=1}^{2^p} d\mathcal{M}_l$. Let $p_t(\boldsymbol{\mathcal{M}})$ be the probability density that the sublattice overlap $\mathcal{M}_l$ takes a value in $(\mathcal{M}_l, \mathcal{M}_l + d\mathcal{M}_l)$ for $l = 1, \cdots, 2^p$ at time t,

$$p_t(\boldsymbol{\mathcal{M}}) = \mathrm{Tr}_{\boldsymbol{s}} \ p_t(\boldsymbol{s}) \prod_{l'=1}^{2^p} \delta(\mathcal{M}_{l'} - \frac{2^p}{N} \sum_{i \in \Lambda_{l'}} s_i). \tag{20}$$

The evolution equation for $\mathcal{M}_l$ is

$$\frac{d\mathcal{M}_l}{dt} = -\mathcal{M}_l + \tanh[\beta h^l(\boldsymbol{\mathcal{M}})],\ l = 1, \cdots, 2^p. \tag{21}$$

Here, $h^l(\boldsymbol{\mathcal{M}})$ is expressed as

$$h^l(\boldsymbol{\mathcal{M}}) = \sum_{\mu=1}^{v} \zeta_\mu \Xi_{\mu,l} \frac{1}{2^p} (\Xi \boldsymbol{\mathcal{M}})_\mu,$$

where Ξ is a $v \times 2^p$ matrix whose (μ, l) element is $\Xi_{\mu,l} = \xi^{\mu,l}$. See Appendix A for the derivation. For $l \leq 2^{p-1}$, $h^{l+2^{p-1}}(\boldsymbol{\mathcal{M}}) = -h^l(\boldsymbol{\mathcal{M}})$ holds because $\Xi_{\mu,l+2^{p-1}} = -\Xi_{\mu,l}$. Thus, we have

$$\frac{d\mathcal{M}_{l+2^{p-1}}}{dt} = -\mathcal{M}_{l+2^{p-1}} - \tanh[\beta h^l(\boldsymbol{\mathcal{M}})], \quad l = 1, \cdots, 2^{p-1}, \tag{22}$$

$$\frac{d\mathcal{M}_l}{dt} = -\mathcal{M}_l + \tanh[\beta h^l(\boldsymbol{\mathcal{M}})], \quad l = 1, \cdots, 2^{p-1}. \tag{23}$$

From these equations, we obtain

$$\frac{d(\mathcal{M}_{l+2^{p-1}} + \mathcal{M}_l)}{dt} = -(\mathcal{M}_{l+2^{p-1}} + \mathcal{M}_l).$$

Its solution is

$$\mathcal{M}_l(t) + \mathcal{M}_{l+2^{p-1}}(t) = e^{-t}[\mathcal{M}_l(0) + \mathcal{M}_{l+2^{p-1}}(0)].$$

As $t \to \infty$, we have

$$\mathcal{M}_{l+2^{p-1}} = -\mathcal{M}_l \text{ for } l \leq 2^{p-1}.$$

Therefore, we assume $\mathcal{M}_{l+2^{p-1}} = -\mathcal{M}_l$ for $l \leq 2^{p-1}$ in equilibrium. Thus, m_μ is expressed as

$$m^\mu = \frac{1}{2^{p-1}} \sum_{l=1}^{2^{p-1}} \xi^{\mu,l} \mathcal{M}_l,\ \mu = 1, \cdots, v. \tag{24}$$

Evolution equations for m^μ are also derived. Let us define the probability density $p_t(\boldsymbol{m})$ of overlaps $\boldsymbol{m} = (m^1, m^2, \cdots, m^v)$ at time t as

$$p_t(\boldsymbol{m}) = \mathrm{Tr}_{\boldsymbol{S}}\ p_t(\boldsymbol{s}) \prod_{\mu=1}^{v} \delta(m^\mu - \frac{1}{N} \sum_j \xi_j^\mu s_j). \tag{25}$$

By using the same recipe as that used to derive Eq. (21), we obtain

$$\frac{d}{dt} m^\mu = -m^\mu + \langle\!\langle \xi^\mu \tanh[\beta(\sum_{\nu=1}^{v} \zeta_\nu m^\nu \xi^\nu)] \rangle\!\rangle,\ \mu = 1, \cdots, v. \tag{26}$$

Any stationary state of the evolution equations given by Eq. (26) satisfies the SPEs (14).

2.5 Relationship between static and dynamic stabilities of equilibrium solution

As is proved in Appendix B, the static and dynamic stabilities of any stationary state are the same, and any stable state of the Hopfield model ($\eta = 0$) continuously changes and is stat-

ically and dynamically stable for sufficiently small unlearning coefficients. Furthermore, it is proved in Appendix B that the stationary solution becomes unstable statically and dynamically at the same time in generic situations that all $\zeta_\mu (\mu = 1, \cdots, v)$ are different. Hereafter, we assume $\zeta_1 = \cdots = \zeta_p = 1$ and $\zeta_{p+1} = \cdots = \zeta_v = -\eta$.

3. Case of $p = 3$

In this section, we study the case of $p = 3$. Sublattices are given as

$$\begin{aligned}
\Lambda_1 &= [i|(\xi_i^1, \xi_i^2, \xi_i^3, \xi_i^4, \xi_i^5, \xi_i^6, \xi_i^7) = (1, 1, 1, 1, 1, 1, -1)], \\
\Lambda_2 &= [i|(\xi_i^1, \xi_i^2, \xi_i^3, \xi_i^4, \xi_i^5, \xi_i^6, \xi_i^7) = (1, 1, -1, 1, 1, -1, 1)], \\
\Lambda_3 &= [i|(\xi_i^1, \xi_i^2, \xi_i^3, \xi_i^4, \xi_i^5, \xi_i^6, \xi_i^7) = (1, -1, 1, 1, -1, 1, 1)], \\
\Lambda_4 &= [i|(\xi_i^1, \xi_i^2, \xi_i^3, \xi_i^4, \xi_i^5, \xi_i^6, \xi_i^7) = (1, -1, -1, -1, 1, 1, 1)], \\
\Lambda_5 &= [i|(\xi_i^1, \xi_i^2, \xi_i^3, \xi_i^4, \xi_i^5, \xi_i^6, \xi_i^7) = (-1, -1, -1, -1, -1, -1, 1)], \\
\Lambda_6 &= [i|(\xi_i^1, \xi_i^2, \xi_i^3, \xi_i^4, \xi_i^5, \xi_i^6, \xi_i^7) = (-1, -1, 1, -1, -1, 1, -1)], \\
\Lambda_7 &= [i|(\xi_i^1, \xi_i^2, \xi_i^3, \xi_i^4, \xi_i^5, \xi_i^6, \xi_i^7) = (-1, 1, -1, -1, 1, -1, -1)], \\
\Lambda_8 &= [i|(\xi_i^1, \xi_i^2, \xi_i^3, \xi_i^4, \xi_i^5, \xi_i^6, \xi_i^7) = (-1, 1, 1, 1, -1, -1, -1)],
\end{aligned} \tag{27}$$

where $\xi_i^4, \xi_i^5, \xi_i^6$, and ξ_i^7 are defined as

$$\xi_i^4 \equiv \xi_i^{(1,2,3:1,1,1)} = \mathrm{sgn}(\xi_i^1 + \xi_i^2 + \xi_i^3), \tag{28}$$

$$\xi_i^5 \equiv \xi_i^{(1,2,3:1,1,-1)} = \mathrm{sgn}(\xi_i^1 + \xi_i^2 - \xi_i^3), \tag{29}$$

$$\xi_i^6 \equiv \xi_i^{(1,2,3:1,-1,1)} = \mathrm{sgn}(\xi_i^1 - \xi_i^2 + \xi_i^3), \tag{30}$$

$$\xi_i^7 \equiv \xi_i^{(1,2,3:1,-1,-1)} = \mathrm{sgn}(\xi_i^1 - \xi_i^2 - \xi_i^3). \tag{31}$$

Correspondingly, m^4, m^5, m^6, and m^7 are defined as

$$m^4 = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{(1,2,3:1,1,1)} s_i(t), \tag{32}$$

$$m^5 = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{(1,2,3:1,1,-1)} s_i(t), \tag{33}$$

$$m^6 = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{(1,2,3:1,-1,1)} s_i(t), \tag{34}$$

$$m^7 = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{(1,2,3:1,-1,-1)} s_i(t). \tag{35}$$

Let Ξ and X be the following 7×8 and 7×4 matrices, respectively, whose (μ, l) component is $\xi^{\mu,l}$.

$$\Xi = (X, -X), \tag{36}$$

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix}. \tag{37}$$

Equation (24) in equilibrium is written as

$$\begin{pmatrix} m^1 & m^2 & m^3 & m^4 & m^5 & m^6 & m^7 \end{pmatrix}^{\mathrm{T}} = \frac{1}{4} X \begin{pmatrix} \mathcal{M}_1 & \mathcal{M}_2 & \mathcal{M}_3 & \mathcal{M}_4 \end{pmatrix}^{\mathrm{T}},$$

where T implies the transpose. Let X_4 be the 4×4 matrix whose (μ, l) element is $\xi^{\mu,l}$ for $\mu = 1, \cdots, 4,\ l = 1, \cdots, 4$. Since $|X_4| = 8$, X has an inverse, that is, $m^1, \cdots, m^4$ and $\mathcal{M}_1, \cdots, \mathcal{M}_4$ have a one-to-one correspondence. Therefore, in the equilibrium state, m^5, m^6, and m^7 are expressed in terms of $m^1, \cdots, m^4$. X_4^{-1} is

$$X_4^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & -1 \\ 0 & 0 & -1 & 1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix}. \tag{38}$$

Thus, we have the following relations in equilibrium:

$$\begin{pmatrix} \mathcal{M}_1 \\ \mathcal{M}_2 \\ \mathcal{M}_3 \\ \mathcal{M}_4 \end{pmatrix} = 4X_4^{-1} \begin{pmatrix} m^1 \\ m^2 \\ m^3 \\ m^4 \end{pmatrix} = 2 \begin{pmatrix} m^1 + m^2 + m^3 - m^4 \\ -m^3 + m^4 \\ -m^2 + m^4 \\ m^1 - m^4 \end{pmatrix}, \tag{39}$$

$$\begin{pmatrix} m^5 \\ m^6 \\ m^7 \end{pmatrix} = \begin{pmatrix} m^1 + m^2 - m^4 \\ m^1 + m^3 - m^4 \\ -m^2 - m^3 + m^4 \end{pmatrix}. \tag{40}$$

From Eq. (19), it follows that the dynamic stability of $\{m^\mu\}$ and that of $\{\mathcal{M}_l\}$ are the same. Also, from Eq. (39), it follows that the static stability of $\{m^\mu\}_{\mu=1,\cdots,4}$ and that of $\{\mathcal{M}_l\}_{l=1,\cdots,4}$ are the same.

3.1 Case that one mixed state is unlearned

In this subsection, we study the case that only $\boldsymbol{\xi}^4 = \boldsymbol{\xi}^{(1,2,3;1,1,1)}$ is unlearned. The SPEs are

$$m^\mu = \langle\langle \xi^\mu \tanh(\beta W_i)\rangle\rangle,\ \mu = 1, \cdots, 4, \tag{41}$$

$$W_i = \sum_{\mu=1}^{3} \xi_i^\mu m^\mu - \eta \xi_i^4 m^4. \tag{42}$$

Here, we put $\zeta_4 = -\eta$. Let us describe the solutions of the SPEs. Let (m^1, m^2, m^3, m^4) be a solution of the SPEs (41). Then, $(m^{1\prime}, m^{2\prime}, m^{3\prime}, m^4)$, in which $m^{1\prime}, m^{2\prime}$, and $m^{3\prime}$ are a permutation of m^1, m^2, and m^3, is also a solution of the SPEs (41). We call this the permutation symmetry. Furthermore, $-\boldsymbol{m}$ is also a solution of the SPEs (41). This is the inversion symmetry. Because of the inversion and permutation symmetries, we assume $m^1 > 0$ and $m^1 \geq m^2 \geq m^3$ without loss of generality. The following types of solution of the SPEs exist.

(1) S_2 state. $m^1 > m^2 = m^3$.

The Hopfield attractor H characterized by $\boldsymbol{m} = (m^1, 0, 0, m^4)$ does not exist when the unlearning term exists. See Appendix C. Since $m^2 = m^3 \simeq 0$ in the S_2 state for $|\eta| \ll 1$, S_2 is regarded as a variant of the Hopfield attractor.

(2) M_4 state. $m^1 = m^2 = m^3$.

This corresponds to the mixed state $\mathrm{M}_4^{(\mathrm{H})}$ in the Hopfield model characterized by $m^1 = m^2 = m^3$.

(3) M_5 state. $m^1 = m^2 \simeq -m^3$.

The mixed state $\mathrm{M}_5^{(\mathrm{H})}$ in the Hopfield model characterized by $m^1 = m^2 = -m^3$ does not exist when the unlearning term exists. See Appendix C. The M_5 state is considered to be a variant of $\mathrm{M}_5^{(\mathrm{H})}$.

(4) S_3 state. m^1, m^2, and m^3 are all different.

(5) P state. $\boldsymbol{m} = (0, 0, 0, 0)$.

In the Hopfield model, there are other mixed states $\mathrm{M}_6^{(\mathrm{H})}$ and $\mathrm{M}_7^{(\mathrm{H})}$ characterized by $m^1 = -m^2 = m^3$ and $m^1 = -m^2 = -m^3$, respectively. In the present model, t other mixed states exist: M_6 characterized by $m^1 = m^3 \simeq -m^2$ corresponding to $\mathrm{M}_6^{(\mathrm{H})}$ and M_7 characterized by $m^1 \simeq -m^2 = -m^3$ corresponding to $\mathrm{M}_7^{(\mathrm{H})}$. Because of the permutation and inversion symmetries, these mixed states M_6 and M_7 have the same static and dynamic stabilities as the M_5 state. Thus, we only have to study the M_4 and M_5 states.

Now, let us show some numerical results. We performed MCMC simulations. The transition probability from $s_k \to -s_k$ is given by Eq. (16).

3.1.1 Statics

As initial conditions, we chose a Hopfield-attractor-like configuration $\boldsymbol{m} = (m^1, 0, 0, m^4)$ and a mixed-state-like configuration $\boldsymbol{m} = (m^1, m^1, m^1, m^4)$. It turned out that N should be of the order 10^5 so that the theoretical and simulation results agree. We took several values of N and several hundred Monte Carlo sweeps (MCSs), where one MCS corresponds to one update of N neurons on average. The overlaps were calculated by averaging the overlaps in the time period from 450 to 500 MCSs. The typical number of samples is 10. In Fig. 1, we display the

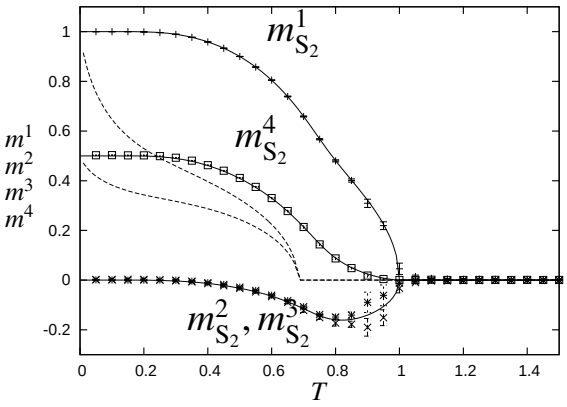

Fig. 1. Temperature dependences of overlaps. $\eta = 0.5$. Curves: theory. Solid curve: S_2, dashed curve: M_4. Symbols: Monte Carlo simulations. $N = 10^5$. Averages are taken from 10 samples. Vertical lines denote the error bars. +: m^1, ×:m^2, *:m^3, □:m^4. Initial configuration: $\boldsymbol{m} = (0.5, 0.5, 0.5, 1)$.

temperature dependences of overlaps for $\eta = 0.5$. In both the Hopfield-attractor-like initial configuration (for which results are not shown) and the mixed-state-like initial configuration, the S_2 state appears for $0 < T < 0.85$ and the S_3 state appears for $0.85 < T \leq 1$. Theoretically, we could not find the S_3 state. Thus, it seems that the theoretical and simulation results do not agree, at least for $0.85 < T \leq 1$. We will discuss this in the next subsection. In Figs. 2(a) and 2(b), we display the temperature dependences of overlaps for $\eta = -0.5$. Starting from a Hopfield-attractor-like initial configuration, the S_2 state appears for $0 < T < 0.57$ and the mixed state M_4 appears for $0.58 < T < 1.42$, as shown in Fig. 2(a). On the other hand, starting from a mixed-state-like initial configuration, the mixed state M_4 appears for $0 < T < 1.42$. See Fig. 2(b). The S_2 state and the unstable S_2 state, which is not drawn in the figure, are annihilated at $T \simeq 0.58$. For $0 < T < 0.57$, the Hopfield attractor and the mixed state M_4 coexist.

The agreement between the numerical and theoretical results is quite good. In the next subsection, we study whether the S_3 state exists for $\eta = 0.5$ at approximately $T = 1$.

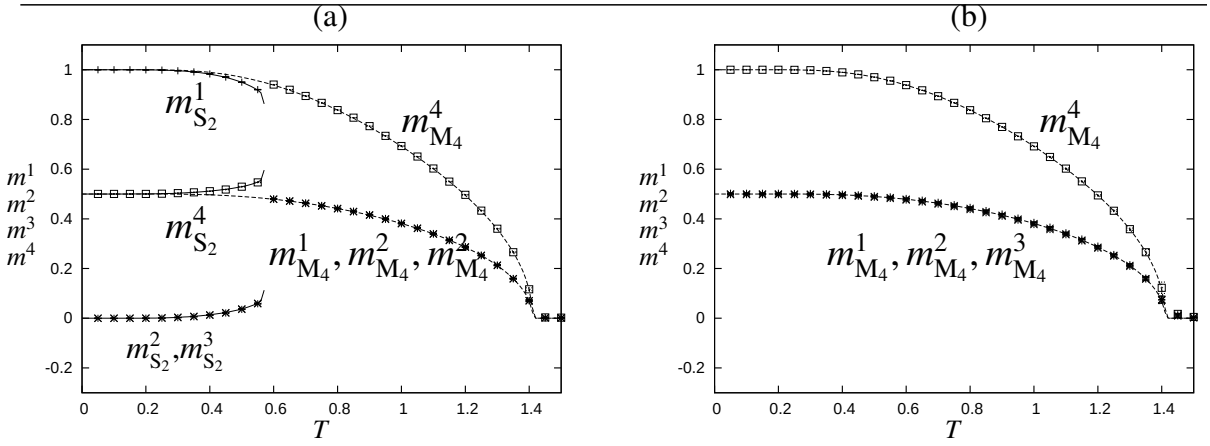

Fig. 2. Temperature dependences of overlaps. $\eta = -0.5$. Curves: theory. Solid curve: S_2, dashed curve: M_4. Symbols: Monte Carlo simulations. $N = 10^5$. Averages are taken from 10 samples. Vertical lines denote the error bars. +: m^1, ×:m^2, *:m^3, □:m^4. (a) Initial configuration: $\boldsymbol{m} = (1, 0, 0, 0.5)$. (b) Initial configuration: $\boldsymbol{m} = (0.5, 0.5, 0.5, 1)$.

3.1.2 Search for S_3 state near critical temperature, $T \simeq T_c$

We performed MCMC simulations for larger values of the system size N and calculated the difference $m^2 - m^3$ at $T = 0.9$ and 0.95. In Fig. 3, for $T = 0.9$, we display $\log_{10} N$ vs $\log_{10}(m^2 - m^3)$ with error bars for $N = 10^5, N = 10^6, N = 3 \times 10^6$, and $N = 10^7$ with numbers of samples of 100, 50, 50, and 20, respectively. We estimated $\log_{10}(m^2 - m^3) \sim$

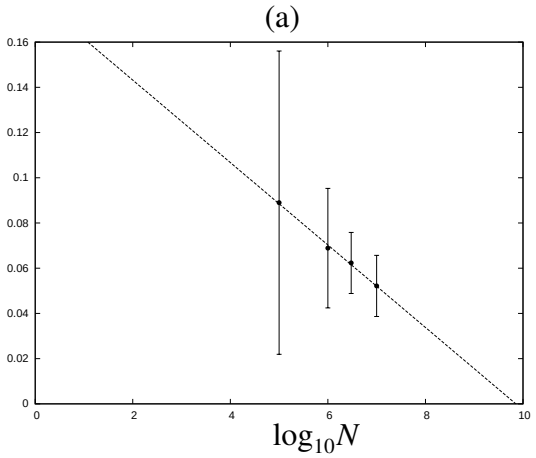

Fig. 3. $\log_{10} N$ vs $\log_{10}(m^2 - m^3)$. $T = 0.9$.

$-0.009926 * \log_{10} N + 0.127675$ at $T = 0.9$ and $\log_{10}(m^2 - m^3) \sim -0.047090 * \log_{10} N + 0.328158$ at $T = 0.95$. From these results, we note that the difference between m^2 and m^3 decreases as N increases. That is, the numerical results of observing the S_3 state is a finite size effect and it is concluded that the S_2 state appears near T_c for $N = \infty$ as the theory predicts.

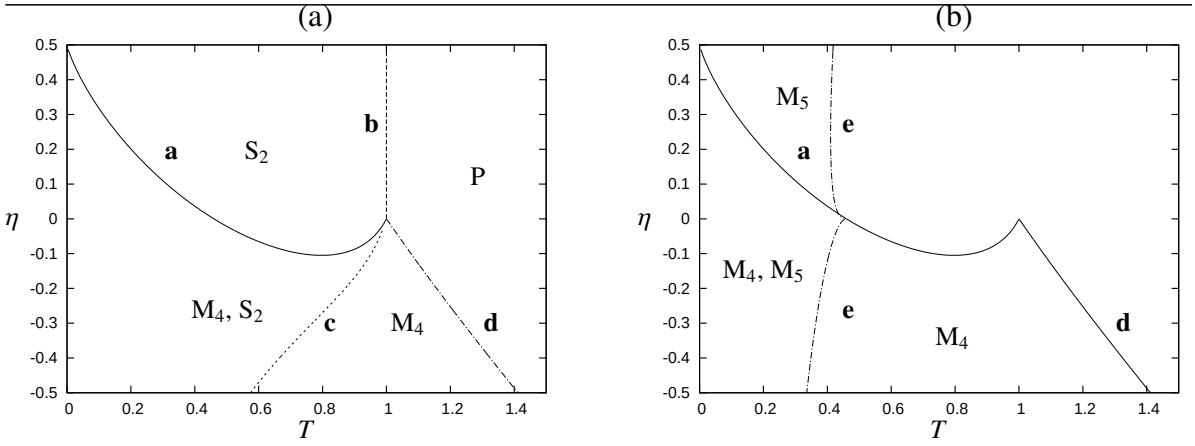

Fig. 4. Phase diagram in (T, η) space determined by using Hessian matrix. (a) Stable regions for P, S_2, and M_4. (b) Stable regions for M_4 and M_5.

3.1.3 Phase diagram

Now, let us study the phase diagram in (T, η) space. The stable regions for solutions are determined by using the Hessian matrix. That is, the boundaries $\boldsymbol{a}, \cdots, \boldsymbol{e}$ in Fig. 4 are determined as follows.

a : Boundary between the S_2 state and the coexistent region of the S_2 and M_4 states. From Fig. 1, we note that at low temperatures, the M_4 state exists but is unstable. The boundary is determined by an eigenvalue of the Hessian matrix for the M_4 state. See Eq. (D·11) in Appendix D.

$$\lambda = \Lambda_{11} - \Lambda_{12} = 0, \text{ i.e., } \frac{\beta}{\cosh^2[\beta(m^1 - \eta m^4)]} = 1.$$

b : Boundary between the P and S_2 states for $\eta > 0$, which is determined by an eigenvalue of the Hessian matrix for the P state. See Eq. (D·7) in Appendix D.

$$\lambda_1 = \lambda_2 = \Lambda_{11} = 1 - \beta = 0, \text{ i.e., } T = 1.$$

c : Boundary between the M_4 state and the coexistent region of the S_2 and M_4 states. This is determined by an eigenvalue of the Hessian matrix for the S_2 state. See Eq. (D·17) in Appendix D. From Fig. 2(a), we note that at this boundary, stable and unstable S_2 states merge and are annihilated.

$$\lambda_4 = 0, \text{ i.e., } -\Lambda_{23}\Lambda_{11}\Lambda_{44} - \Lambda_{11}^2\Lambda_{44} - 4\Lambda_{12}\Lambda_{24}\Lambda_{14} + 2\Lambda_{11}\Lambda_{24}^2 + 2\Lambda_{44}\Lambda_{12}^2 + \Lambda_{14}^2\Lambda_{23} + \Lambda_{11}\Lambda_{14}^2 = 0.$$

d : Boundary between the P and M_4 states for $\eta < 0$, which is determined by the eigenvalue λ_- of the Hessian matrix for the P state. See Eq. (D·8) in Appendix D.

$$\lambda_- = \frac{1}{2}\left(\Lambda_{11} + \Lambda_{44} - \sqrt{(\Lambda_{11} - \Lambda_{44})^2 + 12\Lambda_{14}^2}\right) = 0, \text{ i.e., } \eta = \frac{4T(1-T)}{4T-1}.$$

e : Boundary of the stable M_5 state. The boundary is determined by

$$-\Lambda_{11}\Lambda_{44}\Lambda_{12} - \Lambda_{11}^2\Lambda_{44} - 4\Lambda_{34}\Lambda_{13}\Lambda_{14} + 2\Lambda_{11}\Lambda_{14}^2 + 2\Lambda_{44}\Lambda_{13}^2 + \Lambda_{12}\Lambda_{34}^2 + \Lambda_{11}\Lambda_{34}^2 = 0.$$

See (D·14) in Appendix D.

In Fig. 4, the characters denoted in the regions that are surrounded by the curves represent the stable phases. For $\eta > 0$, by unlearning, the region where the mixed state M_4 is stable decreases in size as η increases, and disappears at $\eta = 0.5$, which is easily proved. On the other hand, the S_2 state exists up to $T = 1$ for $\eta > 0$. Since the S_2 state is regarded as a variant of the Hopfield attractor, it seems that the region where the pattern is stable does not decrease. For $\eta < 0$, as η decreases, the stable region of M_4 increases, but the region where the S_2 state is stable decreases slowly compared with the decrease in the stable region of M_4 for $\eta > 0$ and disappears at about $\eta \sim -1.73$. Figure 4(b) shows the phase diagram only for M_4 and M_5. The stable region of M_5 is almost unchanged for $\eta > 0$ and decreases slowly as $|\eta|$ increases for $\eta < 0$. That is, the unlearning of $\boldsymbol{\xi}^4$ does not cause any significant change in size in the stable region of the mixed state M_5. This is considered to be because $\langle\langle \xi^4 \xi^5 \rangle\rangle = 0$.

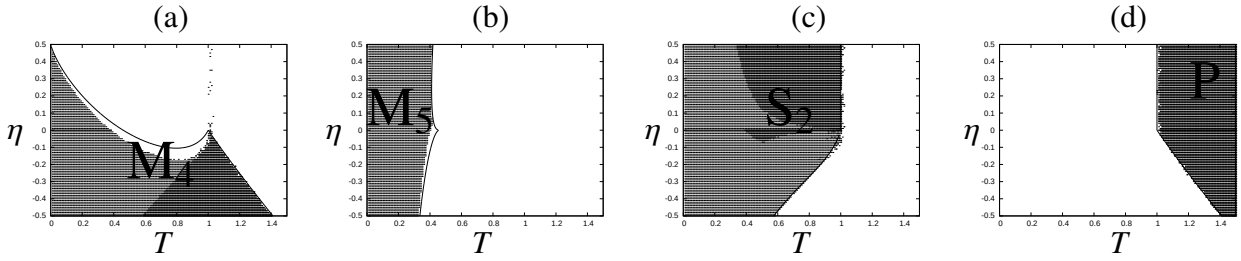

Fig. 5. Stable regions of solutions in (T, η) space. Curves: boundaries obtained by theory using the Hessian matrix, dots: stable regions obtained by MCMC simulation. $N = 10^5$.

To confirm the stable regions for solutions numerically, we performed MCMC simulations and numerically integrated Eq. (26) by the RK method. We iterated the RK routine 10000 times with a time increment of 0.01. This implies that we integrate the evolution equations up to $t = 100$, which is considered to be sufficient for trajectories to converge (as seen later in Fig. 7). On the other hand, we performed MCMC simulations for 10^6 MCSs and used the following criteria to decide the convergent state. First of all, we selected the maximum value of $|m^\mu|$ among $|m^1|, |m^2|$, and $|m^3|$, and if $m^\mu < 0$, we changed the signs of all overlaps. Then, we renumbered the overlaps as $m^1 > m^2 > m^3$. We started with both the Hopfield-attractor-like and mixed-state-M_4-like initial conditions. We successively imposed the following conditions to decide the convergent states.

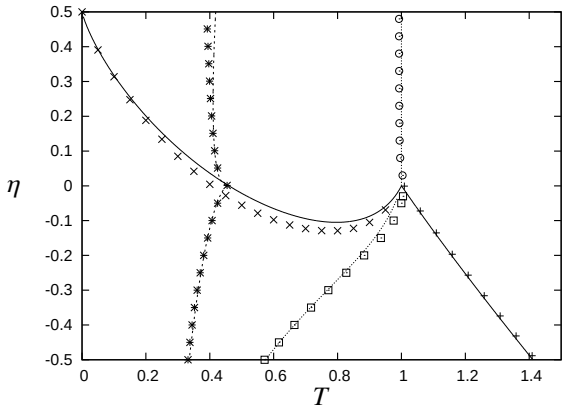

Fig. 6. Phase boundaries obtained by theory using the Hessian matrix (curves) and by the RK method (symbols).

(1) If $m^1 < 0.04$, the state is regarded as P,

(2) else if $m^1 - m^2 < 0.08$ and $m^1 - m^3 < 0.08$, the state is regarded as M_4,

(3) else if $|m^2| < 0.005$ and $|m^3| < 0.005$, the state is regarded as H,

(4) else if $|m^2 - m^3| < 0.01$, the state is regarded as S_2,

(5) else, the state is regarded as S_3.

The results obtained for both initial conditions are used to decide the convergent states. On the other hand, to determine the stable region of M_5, we started the M_5-like initial condition and added the following criterion after (3),

(3') else if $|m^1 - m^2| < 0.25$ and $|m^1 - m^3| < 0.25$, the state is regarded as M_4,

In Fig. 5, we display the theoretical results obtained by using the Hessian matrix and the numerical results obtained by MCMC simulations. In Fig. 6, we draw the phase boundaries obtained using the Hessian matrix and by the RK method. On the boundary of the M_4 state for $T < 1$, the results obtained by MCMC simulations and using the Hessian matrix in Fig. 5(a) do not agree very well. The results obtained by the RK method and using the Hessian matrix in Fig. 6 also do not agree very well. We consider that the reason for this is that the criteria for determining the boundaries of stable regions of solutions by MCMC simulations have ambiguities, that is, we chose conditions such as 0.04 in criterion (1) heuristically. Except for at this boundary, the theoretical and numerical results agree reasonably well.

3.1.4 Dynamics

We performed MCMC simulations for $N = 10^5$ by taking Hopfield-attractor-like and mixed-state-like initial configurations. We compare simulation results obtained by MCMC

simulations with theoretical ones obtained by the RK method. In Fig. 7(a), we display the time series of overlaps for $\eta = 0.5$ and $T = 0.2$. Only the S_2 state exists in this region. Under both initial conditions, all trajectories tend to the S_2 state. In Fig. 7(b), we display the time series of overlaps for $\eta = -0.5$ and $T = 0.2$. In this region, S_2 and M_4 coexist. Depending on the initial conditions, trajectories tend to the S_2 state and the mixed state M_4. The theoretical and simulation results agree quite well.

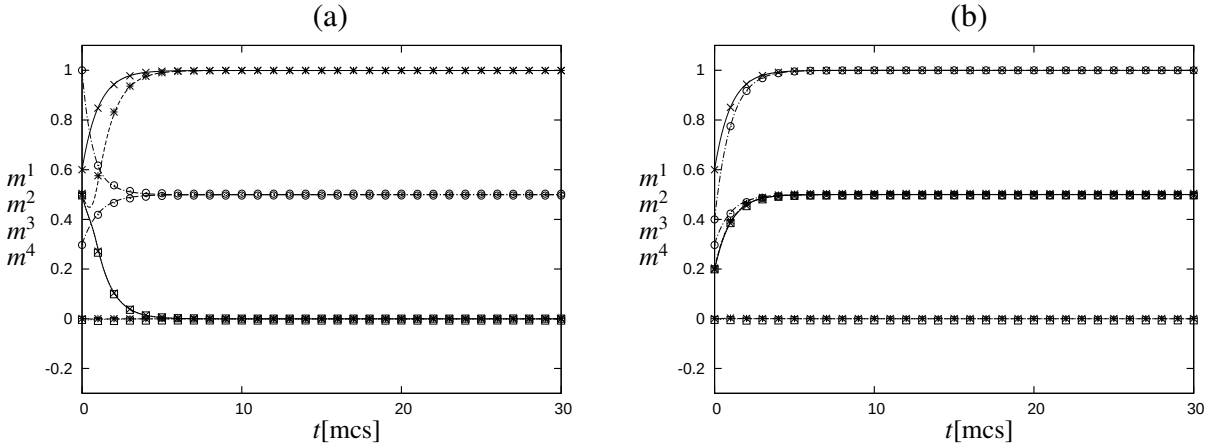

Fig. 7. Curves: theory (RK). Solid curve: m^1, dashed curve: m^2, dotted curve: m^3, dashed-dotted curve: m^4. Symbols: Monte Carlo simulations. $N = 10^5$. $\times$: m^1, $*$: m^2, $\square$: m^3, $\circ$: m^4. (a) $\eta = 0.5$, $T = 0.2$. Initial conditions are $\boldsymbol{m} = (0.6, 0, 0, 0.3)$ and $\boldsymbol{m} = (0.49, 0.5, 0.49, 1)$. (b) $\eta = -0.5$, $T = 0.2$. Initial conditions are $\boldsymbol{m} = (0.6, 0, 0, 0.3)$ and $\boldsymbol{m} = (0.196, 0.204, 0.199, 0.4)$.

The case that another mixed state is unlearned is reduced to the unlearning of $\boldsymbol{\xi}^4$ as follows. Let us consider the unlearning of $\boldsymbol{\xi}^{(1,2,3;\gamma_1,\gamma_2,\gamma_3)}$. Defining quantities as $\boldsymbol{\xi}^{\mu\prime} = \gamma_\mu \boldsymbol{\xi}^\mu, m^{\mu\prime} = \gamma_\mu m^\mu$ ($\mu = 1, 2, 3$), and $\xi_i^{4\prime} = \xi_i^{(1,2,3;\gamma_1,\gamma_2,\gamma_3)} = \mathrm{sgn}(\xi_i^{1\prime} + \xi_i^{2\prime} + \xi_i^{3\prime})$, $m^{4\prime} = m^{(1,2,3;\gamma_1,\gamma_2,\gamma_3)}$, the SPEs become

$$m^{\mu\prime} = \langle\langle \xi^{\mu\prime} \tanh[\beta(\sum_{\nu=1}^{3} \xi^{\nu\prime} m^{\nu\prime} - \eta \xi^{4\prime} m^{4\prime})]\rangle\rangle_{\xi^\prime}. \tag{43}$$

This is simply the SPEs for the unlearning of $\boldsymbol{\xi}^4$. Thus, the stable region of a memory state coincides with one of the stable regions of the memory states for the unlearning of $\boldsymbol{\xi}^4$. The same holds true for mixed states. Thus, we next study the case that multiple mixed states are unlearned.

3.2 Case that multiple mixed states are unlearned

In this subsection, we describe numerical results for the unlearning of multiple mixed states. Hereafter, we use new notations to identify solutions. The memory state corresponding

to $\boldsymbol{\xi}^{\mu}$ and the mixed state corresponding to $\boldsymbol{\xi}^{(\boldsymbol{\mu};\boldsymbol{\gamma})}$ are denoted by M_{μ} and $\mathrm{M}_{(\boldsymbol{\mu};\boldsymbol{\gamma})}$, respectively. In particular, for simplicity, when $\boldsymbol{\gamma} = (1,1,1)$, we use $\boldsymbol{\xi}^{\boldsymbol{\mu}}$ and $\mathrm{M}_{\boldsymbol{\mu}}$ instead of $\boldsymbol{\xi}^{(\boldsymbol{\mu};1,1,1)}$ and $\mathrm{M}_{(\boldsymbol{\mu};1,1,1)}$, respectively. Since the S_2 state is considered to be a variant of the Hopfield attractor, we do not distinguish the M_1 and S_2 states and denote both of them by M_1. We numerically integrated the evolution equations given by Eq. (26) by the RK method starting from memory states or mixed states in order to obtain stable regions of these states. Suppose that we study the overlaps $m^1, m^2, \cdots, m^{\nu}$, and let m^{ω} correspond to the state M_{μ} or $\mathrm{M}_{(\boldsymbol{\mu};\boldsymbol{\gamma})}$. For each (T,η), we set the initial conditions to $m^{\omega} = 0.9$ and $m^{\nu} = \langle\langle \xi^{\omega}\xi^{\nu}\rangle\rangle + r$ for $\nu \neq \omega$, where r is a random number in $[-0.1, 0.1]$. After the trajectory converges, we determine the maximum overlap m^{μ_m} among $m^1, \cdots, m^{\nu}$. We judge that the parameter (T,η) is in the stable region of M_{μ} or $\mathrm{M}_{(\boldsymbol{\mu};\boldsymbol{\gamma})}$ if $\mu_m = \omega$ and $|m^{\mu_m}| > 0.1$. The criterion of the convergence is that $\|\boldsymbol{m}(t+1) - \boldsymbol{m}(t)\| < 10^{-5}$ after a transient, i.e., for $t > 10$. Here, $\|\boldsymbol{m}\| = \sqrt{\sum_{i=1}^{\nu}(m^i)^2}$. This criterion to determine the stable region gives similar results to those in Fig. 5 when $\boldsymbol{\xi}^4$ is unlearned. In Figs. 8-10, we show the results of unlearning two, three, and all mixed states, respectively. We note that the stable regions of memory states are reduced in size for $\eta > 0$ when multiple mixed states are unlearned. The reason for this is considered to be that the mixed states and memory states are correlated. In order to clarify that this is generally true, we study the case that all mixed states are unlearned with equal weight in the next section for a general value of p.

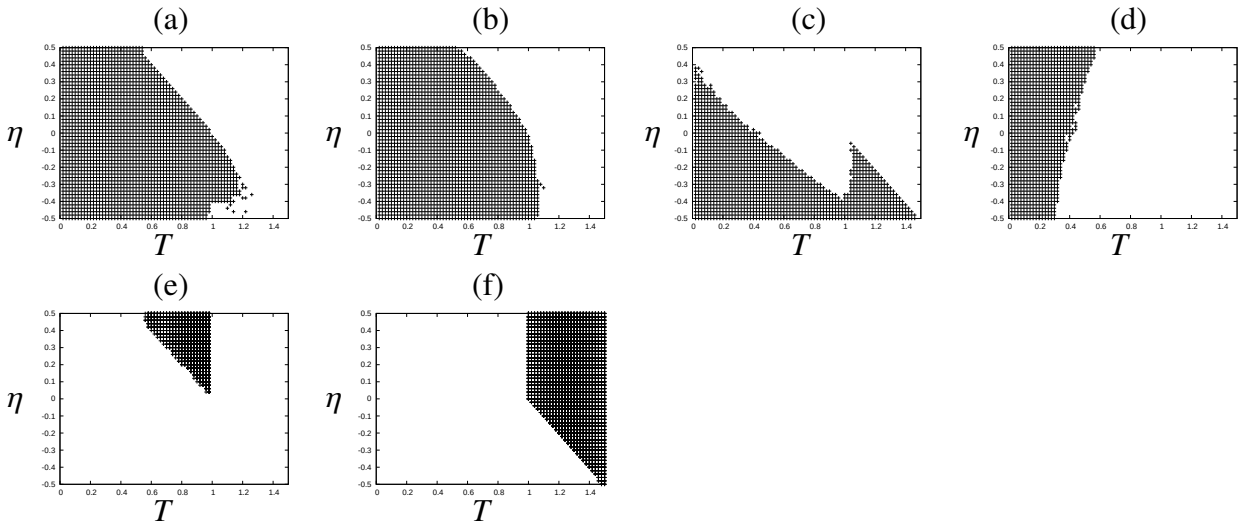

Fig. 8. $p = 3$. $\boldsymbol{\xi}^{(1,2,3)}$ and $\boldsymbol{\xi}^{(1,2,3;1,1,-1)}$ are unlearned with equal weight. Stable regions of solutions are shown in (T,η) space. Dots: stable regions obtained by the RK method. (a) M_1, (b) M_3, (c) $\mathrm{M}_{(1,2,3)}$, (d) $\mathrm{M}_{(1,2,3;1,-1,1)}$, (e) a special state in which $m^1 = -m^2 = m^6, m^3 = m^4 = m^5 = 0$, (f) para state.

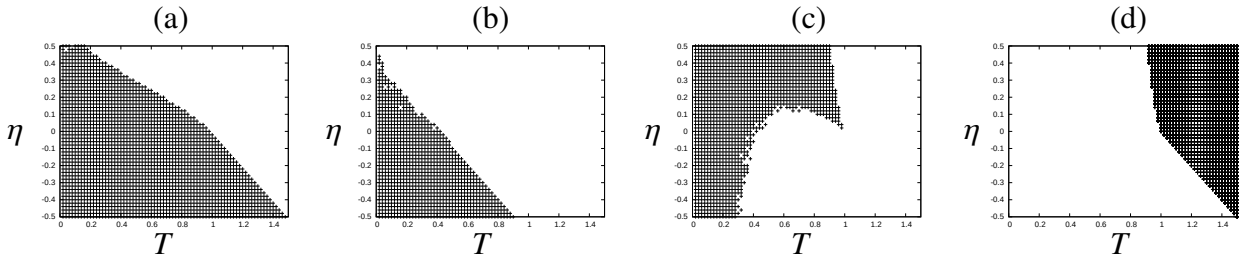

Fig. 9. $p = 3$. Three mixed states $\boldsymbol{\xi}^{(1,2,3)}, \boldsymbol{\xi}^{(1,2,3;1,1,-1)}$, and $\boldsymbol{\xi}^{(1,2,3;1,-1,1)}$ are unlearned with equal weight. Stable regions of solutions are shown in (T, η) space. Dots: stable regions obtained by the RK method. (a) M_1, (b) $\mathrm{M}_{(1,2,3)}$, (c) $\mathrm{M}_{(1,2,3;1,-1,-1)}$, (d) para state.

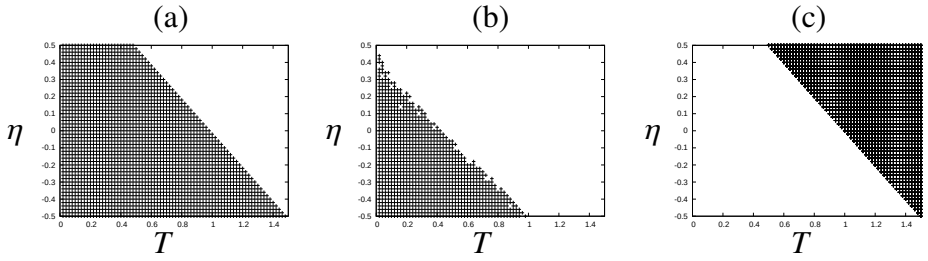

Fig. 10. $p = 3$. All mixed states $\boldsymbol{\xi}^{(1,2,3)}, \boldsymbol{\xi}^{(1,2,3;1,1,-1)}, \boldsymbol{\xi}^{(1,2,3;1,-1,1)}$, and $\boldsymbol{\xi}^{(1,2,3;1,-1,-1)}$ are unlearned with equal weight. Stable regions of solutions are shown in (T, η) space. Dots: stable regions obtained by the RK method. (a) M_1, (b) $\mathrm{M}_{(1,2,3)}$, (c) para state.

4. Case that All Mixed States are Unlearned

For convenience, we include $\gamma_1 = -1$ in the summation of mixed states $\sum_\nu$ in the interaction J_{ij} and rewrite it as

$$J_{ij}^{(\mathrm{U})} = -\frac{\eta}{2N} \sum_{\mathcal{V}'} \xi_i^{(\boldsymbol{\mu};\boldsymbol{\gamma})} \xi_j^{(\boldsymbol{\mu};\boldsymbol{\gamma})}, \tag{44}$$

where $\mathcal{V}' = \mathcal{V} \cup \{(\boldsymbol{\mu}, \boldsymbol{\gamma}) \text{ with } \gamma_1 = -1\}$. The SPEs are

$$m^\mu = \langle\langle \xi^\mu \tanh(\beta W) \rangle\rangle, \ \mu = 1, \cdots, p, \tag{45}$$

$$m^{(\boldsymbol{\mu};\boldsymbol{\gamma})} = \langle\langle \xi^{(\boldsymbol{\mu};\boldsymbol{\gamma})} \tanh(\beta W) \rangle\rangle, \ (\boldsymbol{\mu};\boldsymbol{\gamma}) \in \mathcal{V}', \tag{46}$$

$$W \equiv \sum_{\nu=1}^{p} \xi^\nu m^\nu - \frac{\eta}{2} \sum_{\mathcal{V}'} \xi^{(\boldsymbol{\mu},\boldsymbol{\gamma})} m^{(\boldsymbol{\mu},\boldsymbol{\gamma})}. \tag{47}$$

The following relation is useful for deriving the equations below.

$$\mathrm{sgn}(\xi^1 + \xi^2 + \xi^3) = \frac{1}{2}(\xi^1 + \xi^2 + \xi^3 - \xi^1\xi^2\xi^3). \tag{48}$$

Let us define the critical coefficients $\eta_{\mathrm{c}}^{\mathrm{pattern}}(T)$ and $\eta_{\mathrm{c}}^{\mathrm{mix}}(T)$ at temperature T below which the pattern and the mixed states exist and are stable as solutions of the SPEs, respectively. These coefficients should satisfy $\eta_{\mathrm{c}}^{\mathrm{pattern}}(T_{\mathrm{c}}^{\mathrm{pattern}}) = 0$ and $\eta_{\mathrm{c}}^{\mathrm{mix}}(T_{\mathrm{c}}^{\mathrm{mix}}) = 0$. Here, $T_{\mathrm{c}}^{\mathrm{pattern}} = 1$ and $T_{\mathrm{c}}^{\mathrm{mix}} \simeq 0.46$ are critical temperatures under which the pattern and mixed states with three

patterns exist for the Hopfield model, respectively. In this section, we derive the formulae for $\eta_{\rm c}^{\rm pattern}(0)$ and $\eta_{\rm c}^{\rm mix}(0)$. Since all patterns have the same stability and so do all mixed states because of the permutation and inversion symmetries, we take $\boldsymbol{\xi}^1$ and $\boldsymbol{\xi}^{(\boldsymbol{\mu}_0;\boldsymbol{\gamma}_0)}$ with $\boldsymbol{\mu}_0 = (1,2,3)$ and $\boldsymbol{\gamma}_0 = (1,1,1)$ as a memory state and a mixed state, respectively. Firstly, let us consider $\boldsymbol{\xi}^1$. When $T = 0$, m^1 becomes 1. Thus, we set $s_i = \xi_i^1$. Since $m^\mu = \delta_{\mu,1}$ and $m^{(\boldsymbol{\mu},\boldsymbol{\gamma})} = \langle\langle \xi^{(\boldsymbol{\mu},\boldsymbol{\gamma})}\xi^1\rangle\rangle = \frac{1}{2}\gamma_1\delta_{\mu_1,1}$, W_i is calculated as

$$W_i = \xi_i^1 - \eta \sum_{1<\mu_2<\mu_3} \sum_{\gamma_2} \sum_{\gamma_3} \frac{1}{2}\xi_i^{(\boldsymbol{\mu},\boldsymbol{\gamma})} = \left(1 - \frac{\eta}{2}(p-1)(p-2)\right)\xi_i^1. \tag{49}$$

The SPEs given by Eq. (45) become

$$\begin{aligned}\delta_{\mu,1} &= \langle\langle \xi^\mu \tanh(\beta W)\rangle\rangle = \langle\langle \xi^\mu\xi^1 \tanh[\beta\left(1 - \frac{\eta}{2}(p-1)(p-2)\right)]\rangle\rangle \\ &= \delta_{\mu,1}\tanh[\beta\left(1 - \frac{\eta}{2}(p-1)(p-2)\right)].\end{aligned} \tag{50}$$

Taking the limit $\beta \to \infty$, if $1 - \frac{\eta}{2}(p-1)(p-2) > 0$, the SPEs are satisfied. Thus, we obtain

$$\eta_{\rm c}^{\rm pattern}(0) = \frac{2}{(p-1)(p-2)}. \tag{51}$$

Next, let us study the SPEs given by Eq. (46). They become

$$\frac{1}{2}\gamma_1\delta_{\mu_1,1} = \langle\langle \xi^{(\boldsymbol{\mu};\boldsymbol{\gamma})} \tanh(\beta W)\rangle\rangle = \langle\langle \xi^{(\boldsymbol{\mu};\boldsymbol{\gamma})}\xi^1\rangle\rangle \tanh[\beta\left(1 - \frac{\eta}{2}(p-1)(p-2)\right)]. \tag{52}$$

This is automatically satisfied in the limit $\beta \to \infty$ for $\eta < \eta_{\rm c}^{\rm pattern}(0)$. Now, let us study the stability of $\boldsymbol{\xi}^1$ for $\eta < \eta_{\rm c}^{\rm pattern}(0)$. In this case, $W \neq 0$. The Hessian matrix is

$$\Lambda_{\mu\nu} = \begin{cases} \delta_{\mu\nu} - \beta\langle\langle \xi^\mu\xi^\nu \cosh^{-2}(\beta W)\rangle\rangle, & \mu,\nu \le p, \\ \beta\eta\langle\langle \xi^\mu\xi^\nu \cosh^{-2}(\beta W)\rangle\rangle, & \mu = 1,\cdots,p,\ \nu > p, \\ -\eta\delta_{\mu\nu} - \beta\eta^2\langle\langle \xi^\mu\xi^\nu \cosh^{-2}(\beta W)\rangle\rangle, & \mu,\nu > p. \end{cases} \tag{53}$$

Since $\cosh^{-2}(\beta W) \to 0$ as $\beta \to \infty$, the eigenvalues of Λ are 1 (p-fold) and $-\eta$ (u-fold). Thus, $\boldsymbol{\xi}^1$ is stable for $\eta < \eta_{\rm c}^{\rm pattern}(0)$.

Next, let us study the mixed state $\boldsymbol{\xi}^{(\boldsymbol{\mu}_0,\boldsymbol{\gamma}_0)}$. We set $s_i = \xi_i^{(\boldsymbol{\mu}_0,\boldsymbol{\gamma}_0)}$. W_i is calculated as

$$W_i = \frac{1}{2}(\xi_i^1 + \xi_i^2 + \xi_i^3) - \frac{\eta}{2}\sum_{\boldsymbol{\mu},\boldsymbol{\gamma}} \langle\langle \xi^{(\boldsymbol{\mu},\boldsymbol{\gamma})}\xi^{(\boldsymbol{\mu}_0,\boldsymbol{\gamma}_0)}\rangle\rangle \xi_i^{(\boldsymbol{\mu},\boldsymbol{\gamma})} \tag{54}$$

$$= \frac{1}{2}(\xi_i^1 + \xi_i^2 + \xi_i^3) - \frac{\eta}{2}(W_i^{(1)} + W_i^{(2)} + W_i^{(3)}), \tag{55}$$

where $W_i^{(1)}$, $W_i^{(2)}$, and $W_i^{(3)}$ are the terms in W_i for which one, two, and all of μ_1, μ_2, and μ_3 agree with one, two, and all of 1, 2, and 3, respectively. After a little algebra, we obtain

$$W_i^{(1)} = \frac{(p-3)(p-4)}{2}(\xi_i^1 + \xi_i^2 + \xi_i^3), \tag{56}$$

$$W_i^{(2)} = 2(p-3)(\xi_i^1 + \xi_i^2 + \xi_i^3), \tag{57}$$

$$W_i^{(3)} = 2\mathrm{sgn}(\xi_i^1+\xi_i^2+\xi_i^3) = 2\xi_i^{(\boldsymbol{\mu}_0,\boldsymbol{\gamma}_0)}. \tag{58}$$

Therefore, we obtain

$$\begin{aligned} W_i &= \frac{1}{2}(\xi_i^1+\xi_i^2+\xi_i^3) - \frac{\eta}{4}\Big(p(p-3)(\xi_i^1+\xi_i^2+\xi_i^3) + 4\xi_i^{(\boldsymbol{\mu}_0,\boldsymbol{\gamma}_0)}\Big) \\ &= Vx_i - \eta\xi_i^{(\boldsymbol{\mu}_0,\boldsymbol{\gamma}_0)}, \end{aligned} \tag{59}$$

where $x_i = \xi_i^1+\xi_i^2+\xi_i^3$ and $V = \frac{1}{2} - \frac{\eta}{4}p(p-3)$. In the limit of $\beta\to\infty$, the SPEs given by Eq. (46) become

$$m^{(\boldsymbol{\mu};\boldsymbol{\gamma})} = \langle\langle \xi^{(\boldsymbol{\mu};\boldsymbol{\gamma})}\mathrm{sgn}\Big(Vx_i - \eta\xi^{(\boldsymbol{\mu}_0,\boldsymbol{\gamma}_0)}\Big)\rangle\rangle. \tag{60}$$

If $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ and $\boldsymbol{\gamma}=\boldsymbol{\gamma}_0$, we have

$$1 = \langle\langle \mathrm{sgn}(V|x_i|-\eta)\rangle\rangle. \tag{61}$$

From this, we obtain

$$\eta_{\mathrm{c}}^{\mathrm{mix}}(0) = \frac{2}{(p-1)(p-2)+2}. \tag{62}$$

If none of the elements of $\boldsymbol{\mu}$ agree with the elements of $\boldsymbol{\mu}_0$, both sides of Eq. (60) are 0. Let us consider the case that one of the elements of $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_0$ agree. We assume $\mu_1 = 1$. Then, Eq. (60) becomes

$$\frac{1}{4}\gamma_1 = \langle\langle \mathrm{sgn}(y_i)\ \mathrm{sgn}\Big(Vx_i - \eta\ \mathrm{sgn}(x_i)\Big)\rangle\rangle, \tag{63}$$

where $y_i = \gamma_1\xi_i^1 + \gamma_2\xi_i^{\mu_2} + \gamma_3\xi_i^{\mu_3}$. The right-hand side (r.h.s.) of Eq. (63) is calculated as

$$\langle\langle \mathrm{sgn}(y_i)\ \mathrm{sgn}\Big(Vx_i - \eta\ \mathrm{sgn}(x_i)\Big)\rangle\rangle = \frac{\gamma_1}{8}\Big(\mathrm{sgn}(3V-\eta) + \mathrm{sgn}(V-\eta)\Big). \tag{64}$$

If $\eta < \eta_{\mathrm{c}}^{\mathrm{mix}}(0)$, $V > \eta$ follows. Thus, the r.h.s. of Eq. (63) is equal to $\frac{\gamma_1}{4}$ and to the left-hand side (l.h.s.) of Eq. (63). Next, we study the case that two elements of $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_0$ agree. We assume $\mu_1 = 1$ and $\mu_2 = 2$. Equation (60) becomes

$$\frac{1}{4}(\gamma_1+\gamma_2) = \langle\langle \mathrm{sgn}(y_i)\ \mathrm{sgn}\Big(Vx_i - \eta\ \mathrm{sgn}(x_i)\Big)\rangle\rangle, \tag{65}$$

where $y_i = \gamma_1\xi_i^1 + \gamma_2\xi_i^2 + \gamma_3\xi_i^{\mu_3}$. We obtain

$$\text{r.h.s. of Eq. (65)} = \frac{1}{8}(\gamma_1+\gamma_2)\Big(\mathrm{sgn}(3V-\eta) + \mathrm{sgn}(V-\eta)\Big). \tag{66}$$

This is equal to $\frac{1}{4}(\gamma_1+\gamma_2)$ and to the l.h.s. of Eq. (65) if $\eta < \eta_{\mathrm{c}}^{\mathrm{mix}}(0)$. Finally, let us study the case that $\boldsymbol{\mu}=\boldsymbol{\mu}_0$. Eq. (60) becomes

$$\frac{1}{4}\Big(\sum_{k=1}^{3}\gamma_k + \prod_{k=1}^{3}\gamma_k\Big) = \langle\langle \mathrm{sgn}(y_i)\ \mathrm{sgn}\Big(Vx_i - \eta\ \mathrm{sgn}(x_i)\Big)\rangle\rangle, \tag{67}$$

where $y_i = \gamma_1\xi_i^1 + \gamma_2\xi_i^2 + \gamma_3\xi_i^3$. Then, we obtain

$$\text{r.h.s. of Eq. (67)} = \frac{1}{8}\Big((\gamma_1+\gamma_2+\gamma_3-\gamma_1\gamma_2\gamma_3)\mathrm{sgn}(3V-\eta) + (\gamma_1+\gamma_2+\gamma_3+3\gamma_1\gamma_2\gamma_3)\mathrm{sgn}(V-\eta)\Big). \quad (68)$$

This is equal to $\frac{1}{4}(\gamma_1+\gamma_2+\gamma_3+\gamma_1\gamma_2\gamma_3)$ and to the r.h.s. of Eq. (67) if $\eta < \eta_c^{\rm mix}(0)$. Therefore, all SPEs are satisfied. The stability of $\boldsymbol{\xi}^{(\mu_0;\gamma_0)}$ is the same as that of $\boldsymbol{\xi}^1$ because the eigenvalues of Λ are 1 (p-fold) and $-\eta$ (u-fold). Thus, all memory states for $\eta < \eta_c^{\rm pattern}(0)$ and all mixed states for $\eta < \eta_c^{\rm mix}(0)$ are stable at $T = 0$. From Eqs. (51) and (62), $\eta_c^{\rm pattern}(0) > \eta_c^{\rm mix}(0)$ follows. We numerically integrated the evolution equations given by Eq. (26) with $T = 0$ by the RK method starting from the memory states or mixed states for several values of p in order to obtain the stable regions of these states. The criterion of the convergence is that $\|\boldsymbol{m}(t+1) - \boldsymbol{m}(t)\| < 0.0001$ after a transient, i.e., for $t > 50$. In Fig. 11, we display the p dependences of $\eta_c^{\rm pattern}(0)$ and $\eta_c^{\rm mix}(0)$. The numerical results agree with the theoretical ones reasonably well. As an example, in Fig. 12, for $p = 7$, we display the stable regions for the memory state, mixed state, and para state in the (T, η) plane. Thus, we can eliminate all mixed states

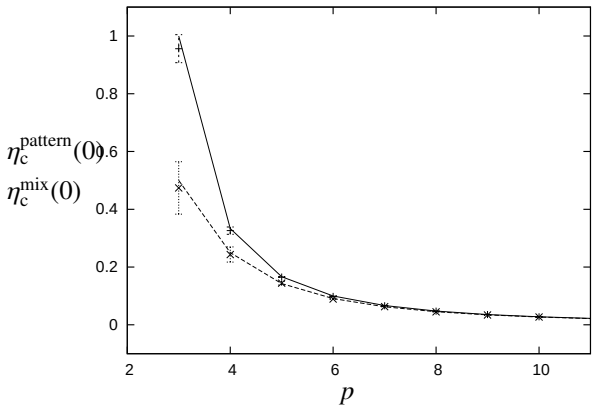

Fig. 11. p dependences of critical unlearning coefficients $\eta_c^{\rm pattern}(0)$ and $\eta_c^{\rm mix}(0)$. Curves: theory. Solid curve: $\eta_c^{\rm pattern}(0)$ and dashed curve: $\eta_c^{\rm mix}(0)$. Symbols: results obtained by the RK method. Averages are taken from about 50 samples. Vertical lines denote the error bars. +: $\eta_c^{\rm pattern}(0)$ and ×: $\eta_c^{\rm mix}(0)$.

and retain all patterns by unlearning in the region $\mathcal{R}$ surrounded by $\eta_c^{\rm mix}(T), \eta_c^{\rm pattern}(T), T = 0$, and $\eta = 0$. In order to tune the parameters (T, η) in this region, there are two opposite factors. Since $\eta_c^{\rm pattern}(0) \to 0$ and $\eta_c^{\rm mix}(0) \to 0$ as $p \to \infty$, the larger the value of p, the more difficult it is to tune the parameters in $\mathcal{R}$. The other factor is the better one for tuning. Because $\eta_c^{\rm pattern}(T_c^{\rm pattern}) = 0$ and $\eta_c^{\rm mix}(T_c^{\rm mix}) = 0$, tuning the parameters in $\mathcal{R}$ seems easier compared with situations in which no such constraints exist.

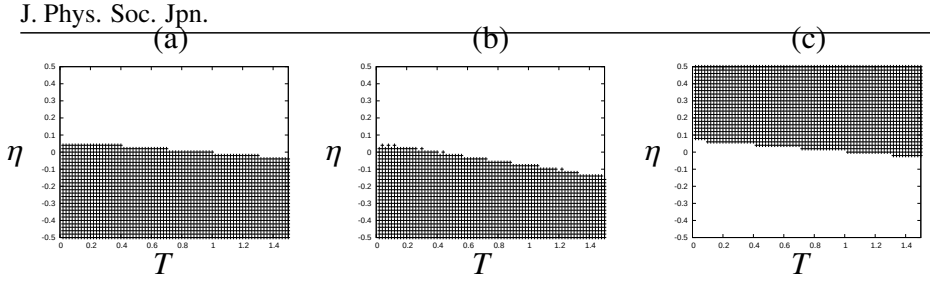

Fig. 12. All mixed states are unlearned. $p = 7$. Stable regions of solutions are shown in (T, η) space obtained by the RK method. (a) M_1, (b) $M_{(1,2,3)}$, (c) para state.

5. Summary and Discussion

We studied unlearning in the Hopfield model for the case of finite loading of patterns in order to remove mixed states and strengthen the stability of memory states. For the finite loading case, mixed states are the only spurious states. As a method of unlearning, we added terms of mixed states with coefficients ζ_μ to the Hamiltonian of the Hopfield model. We set $\zeta_\mu = -\eta$ throughout the paper when we perform numerical calculations. These terms have the effects of removing the mixed states for $\eta > 0$ and learning them for $\eta < 0$.

In this work, firstly, we studied general situations that any number of mixed states is unlearned and obtained the following results. We derived the SPEs and the evolution equations by introducing sublattices and sublattice overlaps. We postulated that the stability at the equilibrium is determined by the condition that the para state is stable at high temperatures. We proved that the static stability determined by this assumption agrees with the dynamic stability in general situations and that any stable state of the Hopfield model ($\eta = 0$) continuously changes and is statically and dynamically stable for sufficiently small unlearning coefficients. Furthermore, we proved that the stationary solution of the SPEs becomes unstable statically and dynamically at the same time in generic situations.

Next, we studied the case of $p = 3$ in detail. In the case that the single mixed state $\boldsymbol{\xi}^4 = \boldsymbol{\xi}^{(1,1,1)}$ is unlearned, we obtained the following results. For $\eta \neq 0$, the Hopfield attractor does not exist; instead, the S_2 state appears. In the S_2 state, $m^1 (> 0)$ is largest and $m_2 = m_3$. This solution is regarded as a variant of the Hopfield attractor. Similarly, the mixed state $M_5^{(H)}$ for $\eta = 0$, which is characterized by $m^1 = m^2 = -m^3$, changes to M_5 with $m^1 = m^2 \neq -m^3$ for $\eta \neq 0$. On the other hand, the mixed state characterized by $m_1 = m_2 = m_3$ exists for $\eta \neq 0$ as well.

As η increases from 0, the region where the mixed state M_4 is stable decreases in size and disappears at $\eta = 0.5$. On the other hand, the stable S_2 region exists up to $T = 1$ for any $\eta (> 0)$, at least for $\eta \leq 0.5$. That is, by unlearning, the spurious state $\boldsymbol{\xi}^4$ is deleted while memory states are retained.

The S_3 state in which m_1, m_2, and m_3 are all different could not be found by numerically solving the SPEs, although in the MCMC simulations, the S_3 solution seemed to exist for small values of N and near the critical temperature T_c with η fixed. It turned out that this phenomenon is a finite size effect because the S_3 state tends to the S_2 state as N becomes large.

In the case of $\eta < 0$, that is, in the case of the learning of the mixed state $\boldsymbol{\xi}^4$, as η decreases, the region where the mixed state M_4 is stable increases in size. On the other hand, the region where the S_2 state is stable decreases slowly compared with the decrease in the stable region of M_4 for $\eta > 0$ and disappears at about $\eta \sim -1.73$. We also studied the stability of other mixed states. It is sufficient to study the stability of M_5 owing to the permutation and inversion symmetries that exist in the system. We found that the stable region of M_5 is almost unchanged for $\eta > 0$ and decreases very slowly in size as $|\eta|$ increases for $\eta < 0$. Therefore, we studied the unlearning of other mixed states, $\boldsymbol{\xi}^{(1,2,3;1,1,-1)}$, $\boldsymbol{\xi}^{(1,2,3;1,-1,1)}$, and $\boldsymbol{\xi}^{(1,2,3;1,-1,-1)}$, in order to delete these mixed states. We found that when multiple mixed states are unlearned, the stable regions of memory states are reduced for $\eta > 0$. The reason for this is considered to be that mixed states and memory states are correlated. In order to clarify whether this is generally true, we studied the case that all mixed states are unlearned with equal weight for a general value of p. We defined the critical coefficients $\eta_c^{\text{pattern}}(T)$ and $\eta_c^{\text{mix}}(T)$ at the temperature T below which the pattern and the mixed states exist and are stable as solutions of the SPEs, respectively. We derived the formulae for $\eta_c^{\text{pattern}}(0)$ and $\eta_c^{\text{mix}}(0)$ and their p dependence was numerically confirmed. Therefore, we can eliminate all mixed states and retain all patterns by the unlearning of all mixed states in the region $\mathcal{R}$ surrounded by $\eta_c^{\text{mix}}(T), \eta_c^{\text{pattern}}(T), T = 0$, and $\eta = 0$. In order to tune the parameters (T, η) in this region, there are two opposite factors. Since $\eta_c^{\text{pattern}}(0) \to 0$ and $\eta_c^{\text{mix}}(0) \to 0$ as $p \to \infty$, the larger the value of p, the more difficult it is to tune the parameters in $\mathcal{R}$. The other factor is the better one for tuning. These coefficients satisfy $\eta_c^{\text{pattern}}(T_c^{\text{pattern}}) = 0$ and $\eta_c^{\text{mix}}(T_c^{\text{mix}}) = 0$, where $T_c^{\text{pattern}} = 1$ and $T_c^{\text{mix}} \simeq 0.46$ are critical temperatures below which the pattern and mixed states with three patterns exist for the Hopfield model, respectively. Owing to these constraints, tuning the parameters in $\mathcal{R}$ seems easier compared with situations in which no such constraints exist.

Appendix A: Evolution Equations for Sublattice Overlaps

In this Appendix, we derive evolution equations for sublattice overlaps $\{\mathcal{M}_l\}$. Let $p_t(\boldsymbol{\mathcal{M}})$ be the probability density of $\boldsymbol{\mathcal{M}} = (\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_2^p)$ at time t,

$$p_t(\boldsymbol{\mathcal{M}}) = \mathrm{Tr}_{\boldsymbol{s}}\ p_t(\boldsymbol{s}) \prod_{l'=1}^{2^p} \delta(\mathcal{M}_{l'} - \frac{2^p}{N} \sum_{i\in\Lambda_{l'}} s_i).$$

Its time evolution is calculated as

$$\begin{aligned}
\frac{\partial}{\partial t} p_t(\boldsymbol{\mathcal{M}}) &= \mathrm{Tr}_{\boldsymbol{s}} \frac{\partial p_t(\boldsymbol{s})}{\partial t} \prod_{l'=1}^{2^p} \delta(\mathcal{M}_{l'} - \frac{2^p}{N} \sum_{i\in\Lambda_{l'}} s_i) \\
&= \mathrm{Tr}_{\boldsymbol{s}} \sum_{l=1}^{2^p} \sum_{k\in\Lambda_l} w_k(\boldsymbol{s}) p_t(\boldsymbol{s}) \Big[\delta\{\mathcal{M}_l - \frac{2^p}{N} \sum_{i\in\Lambda_l} s_i - \frac{2^p}{N}(s'_k - s_k)\} - \delta(\mathcal{M}_l - \frac{2^p}{N} \sum_{i\in\Lambda_l} s_i)\Big] \\
&\quad \times \prod_{l'\neq l} \delta(\mathcal{M}_{l'} - \frac{2^p}{N} \sum_{i\in\Lambda_{l'}} s_i) \\
&= \mathrm{Tr}_{\boldsymbol{s}} \sum_{l=1}^{2^p} \sum_{k\in\Lambda_l} w_k(\boldsymbol{s}) p_t(\boldsymbol{s}) \Big\{\frac{\partial}{\partial \mathcal{M}_l} \delta(\mathcal{M}_l - \frac{2^p}{N} \sum_{i\in\Lambda_l} s_i) \cdot (-\frac{2^p}{N})(s'_k - s_k) \prod_{l'\neq l} \delta(\mathcal{M}_{l'} - \frac{2^p}{N} \sum_{i\in\Lambda_{l'}} s_i)\Big\} \\
&= \mathrm{Tr}_{\boldsymbol{s}} \sum_{l=1}^{2^p} \sum_{k\in\Lambda_l} p_t(\boldsymbol{s})\ \frac{2^p}{N}\ 2w_k(\boldsymbol{s}) s_k\ \frac{\partial}{\partial \mathcal{M}_l} \prod_{l'} \delta(\mathcal{M}_{l'} - \frac{2^p}{N} \sum_{i\in\Lambda_{l'}} s_i), \\
&= \mathrm{Tr}_{\boldsymbol{s}} \sum_{l=1}^{2^p} \frac{\partial}{\partial \mathcal{M}_l} \Big[\{\mathcal{M}_l - \frac{2^p}{N} \sum_{k\in\Lambda_l} \tanh(\beta h_k)\} p_t(\boldsymbol{s}) \prod_{l'} \delta(\mathcal{M}_{l'} - \frac{2^p}{N} \sum_{i\in\Lambda_{l'}} s_i)\Big],
\end{aligned}$$

where $s'_k = -s_k$. For $k \in \Lambda_l$, h_k is calculated as

$$\begin{aligned}
h_k &= \sum_{j(\neq k)} J_{kj} s_j = \sum_{j(\neq k)} \frac{1}{N} \sum_{\mu=1}^{v} \zeta_\mu \xi_k^\mu \xi_j^\mu s_j \simeq \sum_{\mu=1}^{v} \zeta_\mu \xi_k^\mu m^\mu = \sum_{\mu=1}^{v} \zeta_\mu \xi^{\mu,l}\ \frac{1}{2^p} \sum_{l'=1}^{2^p} \xi^{\mu,l'} \mathcal{M}_{l'} \\
&\equiv h^l(\boldsymbol{\mathcal{M}}),
\end{aligned}$$

where $\xi^{\mu,l}$ is the value of ξ_k^μ for k in Λ_l. Therefore, we obtain

$$\begin{aligned}
\sum_{k\in\Lambda_l} \tanh(\beta h_k(\boldsymbol{\mathcal{M}})) &= \frac{N}{2^p} \tanh(\beta h^l(\boldsymbol{\mathcal{M}})), \\
h^l(\boldsymbol{\mathcal{M}}) &= \sum_{\mu=1}^{v} \zeta_\mu \xi^{\mu,l}\ \frac{1}{2^p} \sum_{l'=1}^{2^p} \xi^{\mu,l'} \mathcal{M}_{l'}.
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
\frac{\partial}{\partial t} p_t(\boldsymbol{\mathcal{M}}) &= \sum_l \frac{\partial}{\partial \mathcal{M}_l} \Big[\{\mathcal{M}_l - \tanh(\beta h^l(\boldsymbol{\mathcal{M}}))\} \mathrm{Tr}_{\boldsymbol{s}} p_t(\boldsymbol{s}) \prod_l \delta(\mathcal{M}_l - \frac{2^p}{N} \sum_{i\in\Lambda_l} s_i)\Big] \\
&= \sum_l \frac{\partial}{\partial \mathcal{M}_l} \Big[\{\mathcal{M}_l - \tanh(\beta h^l(\boldsymbol{\mathcal{M}}))\} p_t(\boldsymbol{\mathcal{M}})\Big].
\end{aligned}$$

This implies

$$\frac{d\mathcal{M}_l}{dt} = -\mathcal{M}_l + \tanh[\beta h^l(\boldsymbol{\mathcal{M}})],\ l = 1, \cdots, 2^p. \tag{A·1}$$

$h^l(\boldsymbol{\mathcal{M}})$ is rewritten as

$$h^l(\boldsymbol{\mathcal{M}}) = \sum_{\mu=1}^{v} \zeta_\mu \Xi_{\mu,l} \frac{1}{2^p} (\Xi \boldsymbol{\mathcal{M}})_\mu,$$

where Ξ is the $v \times 2^p$ matrix of which (μ, l) element is $\Xi_{\mu,l} = \xi^{\mu,l}$.

Appendix B: Equivalence of Static and Dynamic Stabilities of Solutions of SPEs

Let us study the stability of the stationary solutions of evolution equations given by Eq. (26). Since the derivatives of f with respect to the overlaps are

$$\frac{\partial f}{\partial m^\mu} = \zeta_\mu m^\mu - \zeta_\mu \langle\!\langle \xi^\mu \tanh[\beta(\sum_{\nu=1}^{v} \zeta_\nu m^\nu \xi^\nu]\rangle\!\rangle,\ \mu = 1, 2, \cdots, v, \tag{B·1}$$

Eq. (26) is rewritten as

$$\frac{dm^\mu}{dt} = -\frac{1}{\zeta_\mu} \frac{\partial f}{\partial m^\mu},\ \mu = 1, 2, \cdots, v. \tag{B·2}$$

Let us define $\boldsymbol{m}$ as

$$\boldsymbol{m} = (m^1, m^2, \cdots, m^v)^{\mathrm{T}}, \tag{B·3}$$

where T implies the transpose. Let $\boldsymbol{m}^*$ be a stationary state of Eq. (26) and $\delta\boldsymbol{m}$ be a deviation from $\boldsymbol{m}^*$,

$$\boldsymbol{m}(t) = \boldsymbol{m}^* + \delta\boldsymbol{m}, \tag{B·4}$$

$$\boldsymbol{m}^* = (m^{1*}, m^{2*}, \cdots, m^{v*})^{\mathrm{T}}, \tag{B·5}$$

$$\delta\boldsymbol{m} = (\delta m^1, \delta m^2, \cdots, \delta m^v)^{\mathrm{T}}. \tag{B·6}$$

Thus, we obtain

$$\frac{d}{dt}\delta m^\mu = -\frac{1}{\zeta_\mu} \frac{\partial f}{\partial m^\mu}\bigg|_{\boldsymbol{m}^*+\delta\boldsymbol{m}} \simeq -\frac{1}{\zeta_\mu} \sum_{\nu=1}^{v} \frac{\partial^2 f}{\partial m^\mu \partial m^\nu} \delta m^\nu,\ \mu = 1, 2, \cdots, v. \tag{B·7}$$

By defining $\Lambda_{\mu\nu} = \dfrac{\partial^2 f}{\partial m^\mu \partial m^\nu}\bigg|_{\boldsymbol{m}=\boldsymbol{m}^*}$, Eq. (B·7) is rewritten as

$$\frac{d}{dt}\delta\boldsymbol{m} = -\hat{\Lambda}\delta\boldsymbol{m}, \tag{B·8}$$

where

$$\hat{\Lambda} = \begin{pmatrix} \Lambda_{11} & \cdots & \Lambda_{1p} & \Lambda_{1p+1} & \cdots & \Lambda_{1v} \\ \Lambda_{21} & \cdots & \Lambda_{2p} & \Lambda_{2p+1} & \cdots & \Lambda_{2v} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Lambda_{p1} & \cdots & \Lambda_{pp} & \Lambda_{pp+1} & \cdots & \Lambda_{pv} \\ \frac{1}{\zeta_{p+1}}\Lambda_{p+11} & \cdots & \frac{1}{\zeta_{p+1}}\Lambda_{p+1p} & \frac{1}{\zeta_{p+1}}\Lambda_{p+1p+1} & \cdots & \frac{1}{\zeta_{p+1}}\Lambda_{p+1v} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\zeta_v}\Lambda_{v1} & \cdots & \frac{1}{\zeta_v}\Lambda_{vp} & \frac{1}{\zeta_v}\Lambda_{vp+1} & \cdots & \frac{1}{\zeta_v}\Lambda_{vv} \end{pmatrix}. \tag{B·9}$$

The solution of Eq. (B·8) is

$$\delta\boldsymbol{m}(t) = e^{-\hat{\Lambda}t}\delta\boldsymbol{m}(0). \tag{B·10}$$

Therefore, the stability condition for $\boldsymbol{m}^*$ is that the real parts of all eigenvalues of $\hat{\Lambda}$ are positive. Let us define a $v \times v$ matrix $G(\{\zeta_\mu\})$ as

$$G(\{\zeta_\mu\}) = G(\{\zeta_\mu\})^{\mathrm{T}} = \begin{pmatrix} \zeta_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \zeta_2 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \ddots & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \zeta_p & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \zeta_{p+1} & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \ldots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & \zeta_v \end{pmatrix}. \tag{B·11}$$

By defining $\delta\tilde{\boldsymbol{m}} = G(\{\sqrt{|\zeta_\mu|}\})\delta\boldsymbol{m}$, from Eq. (B·8), we obtain

$$\frac{d}{dt}\delta\tilde{\boldsymbol{m}} = -\tilde{\Lambda}\delta\tilde{\boldsymbol{m}}, \tag{B·12}$$

$$\tilde{\Lambda} = G(\{\frac{\mathrm{sgn}(\zeta_\mu)}{\sqrt{|\zeta_\mu|}}\})^{\mathrm{T}}\Lambda G(\{\frac{1}{\sqrt{|\zeta_\mu|}}\}). \tag{B·13}$$

If $\zeta_\mu > 0$ for all μ, we obtain

$$\tilde{\Lambda} = G(\{\frac{1}{\sqrt{\zeta_\mu}}\})^{\mathrm{T}}\Lambda G(\{\frac{1}{\sqrt{\zeta_\mu}}\}).$$

Therefore, by Sylvester's law of inertia, the numbers of positive, negative, and zero eigenvalues are the same for $\tilde{\Lambda}$ and Λ. Thus, in this case, the dynamic and static stabilities are the same. On the other hand, if a negative ζ_μ exists, we have to use a different argument. It is necessary to study the Hessian matrix in detail.

Hereafter, the ζ are expressed as follows:

$$\zeta_\mu = 1, \ \ \mu = 1, \cdots, p \tag{B·14}$$

$$\zeta_\mu = -\eta_\mu, \quad \mu = p+1, \cdots, v. \tag{B·15}$$

We assume that all the η_μ take nonzero values. We have the following relation:

$$\hat{\Lambda}(\boldsymbol{\eta}) \equiv \hat{\Lambda}(\boldsymbol{\zeta}) = G(\{-\frac{1}{\eta_\mu}\})\Lambda(\boldsymbol{\eta}), \tag{B·16}$$

where we define

$$\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_v), \tag{B·17}$$

$$\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_v). \tag{B·18}$$

Here, we explicitly write down the $\boldsymbol{\eta}$ dependences of Λ and $\hat{\Lambda}$. Let $\lambda_1(\boldsymbol{\eta}), \lambda_2(\boldsymbol{\eta}), \cdots, \lambda_v(\boldsymbol{\eta})$ be the eigenvalues of $\Lambda(\boldsymbol{\eta})$. Similarly, let $\hat{\lambda}_1(\boldsymbol{\eta}), \hat{\lambda}_2(\boldsymbol{\eta}), \cdots, \hat{\lambda}_v(\boldsymbol{\eta})$ be the eigenvalues of $\hat{\Lambda}(\boldsymbol{\eta})$. Let us study the eigenvalues of $\Lambda(\boldsymbol{\eta})$ for the stationary state $\boldsymbol{m}^*$. The components of $\Lambda(\boldsymbol{\eta})$ are expressed as

$$\Lambda(\boldsymbol{\eta})_{\mu\nu} = \begin{cases} \delta_{\mu\nu} - \beta\langle\langle \xi^\mu \xi^\nu \cosh^{-2} W^* \rangle\rangle, & \mu, \nu \le p, \\ \beta\eta_\nu \langle\langle \xi^\mu \xi^\nu \cosh^{-2} W^* \rangle\rangle, & \mu = 1, \cdots, p, \ \nu > p, \\ -\eta_\mu\delta_{\mu\nu} - \beta\eta_\mu\eta_\nu \langle\langle \xi^\mu \xi^\nu \cosh^{-2} W^* \rangle\rangle, & \mu, \nu > p, \end{cases} \tag{B·19}$$

where $W^* \equiv \beta(\sum_{\nu=1}^{3} m^{\nu *}\xi^\nu - \sum_{\nu=p+1}^{v} \eta_\nu m^{\nu *}\xi^\nu)$. We solve the characteristic equation

$$\left|\Lambda(\boldsymbol{\eta}) - \lambda E_v\right| = 0, \tag{B·20}$$

where E_v is the $v \times v$ unit matrix. Since a solution of Eq. (B·20) is a function of $\boldsymbol{\eta}$, we denote it by $\lambda(\boldsymbol{\eta})$. Then, we have

$$h(\boldsymbol{\eta}) \equiv \left|\Lambda(\boldsymbol{\eta}) - \lambda(\boldsymbol{\eta})E_v\right| = 0. \tag{B·21}$$

Assuming that the $|\eta_\mu|$ are small, we expand $\Lambda(\boldsymbol{\eta})$ and $\lambda(\boldsymbol{\eta})$ as Taylor series,

$$\Lambda(\boldsymbol{\eta}) = \Lambda(\mathbf{0}) + \sum_{\nu(>p)} \frac{\partial\Lambda}{\partial\eta_\nu}(\mathbf{0})\eta_\nu + \cdots,$$

$$\lambda(\boldsymbol{\eta}) = \lambda(\mathbf{0}) + \sum_{\nu(>p)} \frac{\partial\lambda}{\partial\eta_\nu}(\mathbf{0})\eta_\nu + \cdots.$$

$\Lambda(\mathbf{0})$ is expressed as

$$\Lambda(\mathbf{0}) = \begin{bmatrix} \Lambda_p^{(\mathrm{H})} & 0_{p,u} \\ 0_{u,p} & 0_{u,u} \end{bmatrix}, \tag{B·22}$$

where $0_{l,m}$ is an $l \times m$ zero matrix and $\Lambda_p^{(\mathrm{H})}$ is the Hessian matrix for the Hopfield model,

which is a $p \times p$ matrix. It is expressed as

$$\Lambda_p^{(\mathrm{H})} = \begin{bmatrix} 1-\beta\langle\langle \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & -\beta\langle\langle \xi^1\xi^2 \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & \cdots & -\beta\langle\langle \xi^1\xi^p \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle \\ -\beta\langle\langle \xi^2\xi^1 \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & 1-\beta\langle\langle \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & \cdots & -\beta\langle\langle \xi^2\xi^p \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle \\ \cdots & \cdots & \cdots & \cdots \\ -\beta\langle\langle \xi^p\xi^1 \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & -\beta\langle\langle \xi^p\xi^2 \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & \cdots & 1-\beta\langle\langle \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle \end{bmatrix} \quad \text{(B·23)}$$

where $W^{(\mathrm{H})} = \beta \sum_{\nu=1}^{p} m^{\nu(\mathrm{H})}\xi^\nu$ and $m^{\nu(\mathrm{H})} (\nu = 1, \cdots, p)$ is the stationary state of the Hopfield model. Now, let us consider the situation that all the unlearning coefficients are of the same order and assume $\eta_\mu = \bar{\eta}_\mu \epsilon$ $(\mu = p+1, \cdots, v)$, where ϵ is a small parameter, $\bar{\eta}_\mu = O(\epsilon^0)$, and $\bar{\eta}_\mu > 0$. The term of $O(\epsilon^0)$ in Eq. (B·21) is $h(\mathbf{0})$, which becomes

$$h(\mathbf{0}) = \left|\Lambda(\mathbf{0}) - \lambda(\mathbf{0})E_v\right| = \begin{vmatrix} \Lambda_p^{(\mathrm{H})} - \lambda(\mathbf{0})E_p & 0_{p,u} \\ 0_{u,p} & -\lambda(\mathbf{0})E_{u,u} \end{vmatrix} \quad \text{(B·24)}$$

$$= \left|\Lambda_p^{(\mathrm{H})} - \lambda(0)E_3\right|(-\lambda(\mathbf{0}))^u = 0. \quad \text{(B·25)}$$

The solutions of $\left|\Lambda_p^{(\mathrm{H})} - \lambda(\mathbf{0})E_p\right| = 0$ are the eigenvalues of the stationary state for the Hopfield model, and we denote them by $\lambda_i^{(\mathrm{H})}$ $(i = 1, \cdots, p)$. Thus, we have $\lambda_i(\mathbf{0}) = \lambda_i^{(\mathrm{H})}$ $(i = 1, \cdots, p)$. We denote other solutions as $\lambda_\nu(\mathbf{0}) = 0$ $(\nu = p+1, \cdots, v)$. If the stationary state of the Hopfield model is stable, $\lambda_i^{(\mathrm{H})} > 0$ $(i = 1, \cdots, p)$. Thus, let us consider $\lambda_\nu(\eta)$ $(\nu > p)$. Let us define $h_\nu(\boldsymbol{\eta})$ as $h(\boldsymbol{\eta})$ evaluated at $\lambda = \lambda_\nu(\eta)$. Since the components of the $(p+1)$th to the vth rows of $\Lambda(\boldsymbol{\eta})$ are of order ϵ or of order higher than ϵ, the lowest order of $h_\nu(\boldsymbol{\eta})$ is $O(\epsilon^u)$ and it is given by $\frac{\partial^u h_\nu}{\partial\eta_{p+1}\cdots\partial\eta_v}(\mathbf{0})\bar{\eta}_{p+1}\cdots\bar{\eta}_v\epsilon^u$. $\frac{\partial^u h_\nu}{\partial\eta_{p+1}\cdots\partial\eta_v}(\mathbf{0})$ is calculated as

$$\frac{\partial^u h_\nu}{\partial\eta_{p+1}\cdots\partial\eta_v}(\mathbf{0}) = \quad \text{(B·26)}$$

$$\begin{vmatrix} & & \Lambda_p^{(\mathrm{H})} & & & & 0_{p,u} \\ \beta\langle\langle \xi^{p+1}\xi^1 \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & \cdots & \beta\langle\langle \xi^{p+1}\xi^p \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & -1-\frac{\partial\lambda_\nu}{\partial\eta_{p+1}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \cdots & 0 \\ \beta\langle\langle \xi^\nu\xi^1 \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & \cdots & \beta\langle\langle \xi^\nu\xi^p \frac{1}{\cosh^2 W^{(\mathrm{H})}}\rangle\rangle & 0 & \cdots & 0 & -1-\frac{\partial\lambda_\nu}{\partial\eta_\nu} \end{vmatrix}$$

$$= \left|\Lambda_p^{(\mathrm{H})}\right| \prod_{\mu=p+1}^{v} (-1 - \frac{\partial\lambda_\nu}{\partial\eta_\mu}). \quad \text{(B·27)}$$

Since $\lambda_i^{(\mathrm{H})} > 0$ $(i = 1, \cdots, p)$, $\left|\Lambda_p^{(\mathrm{H})}\right| \neq 0$ follows. Thus, we have

$$\frac{\partial\lambda_\nu}{\partial\eta_\mu}(\mathbf{0}) = -1, \ \ \mu = p+1, \cdots, v. \quad \text{(B·28)}$$

Therefore, up to $O(\epsilon)$, $\lambda_\nu(\boldsymbol{\eta})$ is expressed as

$$\lambda_\nu(\boldsymbol{\eta}) \simeq \lambda_\nu(\mathbf{0}) + \sum_{\mu=p+1}^{\nu} \frac{\partial \lambda_\nu}{\partial \eta_\mu}(\mathbf{0})\eta_\mu = -\sum_{\mu=p+1}^{\nu} \eta_\mu. \tag{B·29}$$

Next, let us calculate $\hat{\lambda}_i(\mathbf{0})$. We rewrite $\hat{\Lambda}(\boldsymbol{\eta})$ as

$$\hat{\Lambda}(\eta) = \begin{pmatrix} \Lambda_p(\boldsymbol{\eta}) & B(\boldsymbol{\eta}) \\ C(\boldsymbol{\eta}) & D(\boldsymbol{\eta}) \end{pmatrix}, \tag{B·30}$$

where Λ_p, B, C, D, and G are defined as

$$\Big(\Lambda_p(\boldsymbol{\eta})\Big)_{\mu\nu} = \Lambda_{\mu\nu}(\boldsymbol{\eta}),\ \mu,\nu = 1,\cdots,p, \tag{B·31}$$

$$B(\boldsymbol{\eta}) = \begin{pmatrix} \Lambda_{1p+1}(\boldsymbol{\eta}) & \cdots & \Lambda_{1\nu}(\boldsymbol{\eta}) \\ \cdots & \cdots & \cdots \\ \Lambda_{pp+1}(\boldsymbol{\eta}) & \cdots & \Lambda_{p\nu}(\boldsymbol{\eta}) \end{pmatrix},\ p \times u \text{ matrix}, \tag{B·32}$$

$$C(\boldsymbol{\eta}) = G_u(\{-\frac{1}{\eta_\nu}\}) \begin{pmatrix} \Lambda_{p+11}(\boldsymbol{\eta}) & \cdots & \Lambda_{p+1p}(\boldsymbol{\eta}) \\ \cdots & \cdots & \cdots \\ \Lambda_{\nu 1}(\boldsymbol{\eta}) & \cdots & \Lambda_{\nu p}(\boldsymbol{\eta}) \end{pmatrix},\ u \times p \text{ matrix}, \tag{B·33}$$

$$D(\boldsymbol{\eta}) = G_u(\{-\frac{1}{\eta_\nu}\}) \begin{pmatrix} \Lambda_{p+1p+1}(\boldsymbol{\eta}) & \cdots & \Lambda_{p+1\nu}(\boldsymbol{\eta}) \\ \cdots & \cdots & \cdots \\ \Lambda_{\nu p+1}(\boldsymbol{\eta}) & \cdots & \Lambda_{\nu\nu}(\boldsymbol{\eta}) \end{pmatrix},\ u \times u \text{ matrix}, \tag{B·34}$$

$$G_u(\{-\frac{1}{\eta_\nu}\}) = \begin{pmatrix} -\frac{1}{\eta_{p+1}} & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & 0 \\ 0 & 0 & \cdots & -\frac{1}{\eta_\nu} \end{pmatrix},\ u \times u \text{ matrix}. \tag{B·35}$$

Thus, we have

$$\hat{\Lambda}(\mathbf{0}) = \begin{pmatrix} \Lambda_p^{(\mathrm{H})} & 0_{pu} \\ C(\mathbf{0}) & E_u \end{pmatrix}. \tag{B·36}$$

Now, let us solve the characteristic equation of $\hat{\Lambda}$ for the stationary state studied above,

$$\hat{h}(\boldsymbol{\eta}) \equiv \left|\hat{\Lambda}(\boldsymbol{\eta}) - \hat{\lambda}(\boldsymbol{\eta})E_\nu\right| = 0. \tag{B·37}$$

Let us expand $\hat{\Lambda}(\boldsymbol{\eta})$ and $\hat{\lambda}(\boldsymbol{\eta})$ as Taylor series,

$$\hat{\Lambda}(\boldsymbol{\eta}) = \hat{\Lambda}(\mathbf{0}) + \sum_{\mu>p} \frac{\partial \hat{\Lambda}}{\partial \eta_\mu}(\mathbf{0})\eta_\mu + \cdots, \tag{B·38}$$

$$\hat{\lambda}(\boldsymbol{\eta}) = \hat{\lambda}(\mathbf{0}) + \sum_{\mu>p} \frac{\partial \hat{\lambda}}{\partial \eta_\mu}(\mathbf{0})\eta_\mu + \cdots. \tag{B·39}$$

The term of $O(\epsilon^0)$ for Eq. (B·37) is

$$\hat{h}(\mathbf{0}) = \begin{vmatrix} \Lambda_p^{(\mathrm{H})} - \hat{\lambda}(\mathbf{0})E_p & 0_{p,u} \\ C(\mathbf{0}) & (1-\hat{\lambda}(\mathbf{0}))E_u \end{vmatrix} = 0. \tag{B·40}$$

From this, we obtain

$$\left|\Lambda_p^{(\mathrm{H})} - \hat{\lambda}(\mathbf{0})E_p\right|(1-\hat{\lambda}(\mathbf{0}))^u = 0. \tag{B·41}$$

Therefore, we have

$$\left|\Lambda_p^{(\mathrm{H})} - \hat{\lambda}(\mathbf{0})E_p\right| = 0, \tag{B·42}$$

$$1-\hat{\lambda}(\mathbf{0}) = 0. \tag{B·43}$$

The first equation is the characteristic equation for the Hopfield model, and we have

$$\hat{\lambda}_i(\mathbf{0}) = \lambda_i^H > 0,\ i = 1, \cdots, p. \tag{B·44}$$

The second equation gives

$$\hat{\lambda}_\nu(\mathbf{0}) = 1,\ \nu = p+1, \cdots, v. \tag{B·45}$$

Therefore, for sufficiently small $|\epsilon|$, i.e., $|\epsilon| \ll 1$, we have

$$\hat{\lambda}_i(\boldsymbol{\eta}) > 0,\ i = 1, \cdots, v, \tag{B·46}$$

and the stationary state is dynamically stable.

Therefore, for sufficiently small ϵ, we have

$$\begin{aligned} \hat{\lambda}_i(\boldsymbol{\eta}) &> 0\ i = 1, \cdots, v \\ \lambda_i(\boldsymbol{\eta}) &> 0\ i = 1, \cdots, p \\ \lambda_\nu(\boldsymbol{\eta}) &\simeq -\sum_{\mu=p+1}^{v} \eta_\mu,\ i = p+1, \cdots, v. \end{aligned} \tag{B·47}$$

Thus, we conclude that irrespective of the sign of η_μ, any stable stationary state of the Hopfield model is statically and dynamically stable for sufficiently small $|\epsilon|$. In particular, this holds for $\eta_\mu = \eta$ for all μ.

Let us consider the breaking of stability when one of the η, say η_i, changes. For any nonzero ϵ, it follows from Eq. (B·16) that

$$\prod_{i=1}^{v} \hat{\lambda}_i(\boldsymbol{\eta}) = [\prod_{j=p+1}^{v} (-\frac{1}{\eta_i})] \prod_{i=1}^{v} \lambda_i(\boldsymbol{\eta}). \tag{B·48}$$

Let $\eta_c(T)$ be the value of η_i where the stationary state becomes statically unstable for the first time when $|\eta_i|$ is increased from 0 at a fixed temperature T. Similarly, let $\hat{\eta}_c(T)$ be the value

of η_i where the stationary state becomes dynamically unstable for the first time when $|\eta_i|$ is increased from 0 at a fixed temperature T. Generically, only one of the λ, say λ_j, changes its sign. Therefore, at $\eta_c(T)$, $\lambda_j = 0$. Here, we separately discuss the cases $\eta_i < 0$ and $\eta_i > 0$. For the negative η_i, by Sylvester's law of inertia, only $\hat{\lambda}_j$ becomes 0 at $\eta_i = \eta_c(T)$. Thus, $\eta_c(T) = \hat{\eta}_c(T)$ follows. On the other hand, for the positive η_i, from Eq. (B·48), generically, only one of the $\hat{\lambda}$, say $\hat{\lambda}_k$, changes its sign at $\eta_i = \hat{\eta}_c(T)$, and then $\eta_c(T) = \hat{\eta}_c(T)$ follows. Therefore, the stationary solution becomes unstable statically and dynamically at the same time for positive and negative η_i in generic situations. This property is proved to hold when $\eta_\mu = \eta$ for all μ by a similar argument.

In general, we cannot say anything about whether this unstable solution becomes stable when η_i is further increased because eigenvalues other than λ_j, $\hat{\lambda}_j$, and $\hat{\lambda}_k$ may change their signs.

Appendix C: SPEs in the Case that the Single Mixed State is Unlearned

The SPEs for the solutions are given as follows.

(1) Hopfield attractor H. $m^2 = m^3 = 0$.

$$m^1 = \frac{1}{4}\Big(3\tanh[\beta(m^1 - \eta m^4)] + \tanh[\beta(m^1 + \eta m^4)]\Big), \tag{C·1}$$

$$m^2 = \frac{1}{4}\Big(\tanh[\beta(m^1 - \eta m^4)] - \tanh[\beta(m^1 + \eta m^4)]\Big), \tag{C·2}$$

$$m^4 = \frac{1}{4}\Big(3\tanh[\beta(m^1 - \eta m^4)] - \tanh[\beta(m^1 + \eta m^4)]\Big). \tag{C·3}$$

From the condition of $m^2 = 0$, $\eta = 0$ follows. That is, the Hopfield attractor does not exist when the unlearning term exists in the Hamiltonian.

(2) Mixed state M_4, $m_1 = m_2 = m_3$.

$$m^1 = \frac{1}{4}\Big(\tanh[\beta(3m^1 - \eta m^4)] + \tanh[\beta(m^1 - \eta m^4)]\Big), \tag{C·4}$$

$$m^4 = \frac{1}{4}\Big(\tanh[\beta(3m^1 - \eta m^4)] + 3\tanh[\beta(m^1 - \eta m^4)]\Big). \tag{C·5}$$

(3) Mixed state M_5. $m_1 = m_2 (\simeq -m_3)$.

$$m^1 = \frac{1}{4}\Big(\tanh[\beta(2m^1 + m^3 - \eta m^4)] + \tanh[\beta(2m^1 - m^3 - \eta m^4)]\Big), \tag{C·6}$$

$$m^3 = \frac{1}{4}\Big(\tanh[\beta(2m^1 + m^3 - \eta m^4)] - \tanh[\beta(2m^1 - m^3 - \eta m^4)] + 2\tanh[\beta(m^3 - \eta m^4)]\Big), \tag{C·7}$$

$$m^4 = \frac{1}{4}\Big(\tanh[\beta(2m^1 + m^3 - \eta m^4)] + \tanh[\beta(2m^1 - m^3 - \eta m^4)] + 2\tanh[\beta(m^3 - \eta m^4)]\Big). \quad \text{(C·8)}$$

From these equations, we note that $m^3 \neq -m^1$ unless $\eta = 0$.

(4) S_2 state.

$$m^1 = \frac{1}{4}\Big(\tanh[\beta(m^1 + 2m^2 - \eta m^4)] + 2\tanh[\beta(m^1 - \eta m^4)] + \tanh[\beta(m^1 - 2m^2 + \eta m^4)]\Big), \quad \text{(C·9)}$$

$$m^2 = \frac{1}{4}\Big(\tanh[\beta(m^1 + 2m^2 - \eta m^4)] - \tanh[\beta(m^1 - 2m^2 + \eta m^4)]\Big), \quad \text{(C·10)}$$

$$m^4 = \frac{1}{4}\Big(\tanh[\beta(m^1 + 2m^2 - \eta m^4)] + 2\tanh[\beta(m^1 - \eta m^4)] - \tanh[\beta(m^1 - 2m^2 + \eta m^4)]\Big). \quad \text{(C·11)}$$

Appendix D: Stability Analysis of Solutions of SPEs When One Mixed State is Unlearned

In the general case that f is a function of seven overlaps $m^1, \cdots, m^7$, we put $\Lambda_{\mu\nu} = \dfrac{\partial^2 f}{\partial m^\mu \partial m^\nu}$. For $\mu = 1, \cdots, 7$ and $\nu = 1, \cdots, 7$, we obtain

$$\Lambda_{\mu\nu} = \frac{\partial^2 f}{\partial m^\mu \partial m^\nu} = \zeta_\mu\Big(\delta_{\mu\nu} - \beta\zeta_\nu \langle\langle \xi^\mu \xi^\nu \frac{1}{\cosh^2(\beta \sum_{\tau=1}^{7} \zeta_\tau m^\tau \xi^\tau)} \rangle\rangle\Big). \quad \text{(D·1)}$$

The characteristic equation of $\Lambda \equiv \{\Lambda_{\mu\nu}\}$ is $|\Lambda - \lambda E| = 0$. The boundary of a stable region for any solution of the SPEs is determined by $|\Lambda| = 0$ because some eigenvalue becomes 0 at the boundary. When only $\boldsymbol{\xi}^4$ is unlearned, we put $\zeta_1 = \zeta_2 = \zeta_3 = 1$, $\zeta_4 = -\eta$, and $\zeta_5 = \zeta_6 = \zeta_7 = 0$. In this case, the elements of the Hessian matrix are

$$\Lambda_{\mu\nu} = \delta_{\mu\nu} - \beta\langle\!\langle \xi^\mu \xi^\nu \frac{1}{\cosh^2[\beta(\sum_{\mu=1}^{3} m^\mu \xi^\mu - \eta\, m^4 \xi^4)]} \rangle\!\rangle, \quad \text{for } \mu, \nu = 1, 2, 3, \quad \text{(D·2)}$$

$$\Lambda_{\mu 4} = \beta\eta\langle\!\langle \xi^\mu \xi^4 \frac{1}{\cosh^2[\beta(\sum_{\mu=1}^{3} m^\mu \xi^\mu - \eta\, m^4 \xi^4)]} \rangle\!\rangle, \quad \text{for } \mu = 1, 2, 3, \quad \text{(D·3)}$$

$$\Lambda_{44} = -\eta - \beta\eta^2\langle\!\langle \frac{1}{\cosh^2[\beta(\sum_{\mu=1}^{3} m^\mu \xi^\mu - \eta\, m^4 \xi^4)]} \rangle\!\rangle. \quad \text{(D·4)}$$

We below derive characteristic equations for the solutions of the SPEs.

(1) Hopfield attractor. We have

$$\Lambda_{11} = \Lambda_{22} = \Lambda_{33}, \Lambda_{12} = \Lambda_{13}, \Lambda_{24} = \Lambda_{34}.$$

The characteristic equation is

$$\begin{aligned}|\Lambda - \lambda E| &= \{\lambda - (\Lambda_{11} + \Lambda_{12})\}\{\lambda^3 + (\Lambda_{12} - 2\Lambda_{11} - \Lambda_{44})\lambda^2 \\ &+(\Lambda_{11}^2 - \Lambda_{11}\Lambda_{12} - \Lambda 44\Lambda_{12} + 2\Lambda_{11}\Lambda_{44} - 2\Lambda_{12}^2 - 2\Lambda_{24}^2 - \Lambda_{14}^2)\lambda \\ &+(\Lambda_{12}\Lambda_{11}\Lambda 44 - \Lambda_{11}^2\Lambda 44 - 4\Lambda_{12}\Lambda_{24}\Lambda_{14} + 2\Lambda_{12}^2\Lambda 44 + 2\Lambda_{24}^2\Lambda_{11} - \Lambda_{14}^2\Lambda_{12} + \Lambda_{14}^2\Lambda_{11})\} = 0.\end{aligned} \tag{D·5}$$

(2) P state. We have

$$\Lambda_{11} = \Lambda_{22} = \Lambda_{33} = 1-\beta, \Lambda_{44} = -\eta - \beta\eta^2, \Lambda_{12} = \Lambda_{13} = \Lambda_{23} = 0, \Lambda_{14} = \Lambda_{24} = \Lambda_{34} = \frac{1}{2}\beta\eta.$$

Thus,

$$|\Lambda - \lambda E| = (\lambda - \Lambda_{11})^2\{\lambda^2 + (-\Lambda_{11} - \Lambda_{44})\lambda + \Lambda_{11}\Lambda_{44} - 3\Lambda_{14}^2\} = 0. \tag{D·6}$$

The solution of Eq. (D·6) is

$$\lambda_1 = \lambda_2 = \Lambda_{11}, \tag{D·7}$$

$$\lambda_\pm = \frac{1}{2}\left(\Lambda_{11} + \Lambda_{44} \pm \sqrt{(\Lambda_{11} - \Lambda_{44})^2 + 12\Lambda_{14}^2}\right). \tag{D·8}$$

The stable region for the P state is determined by the conditions that $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$ and $\lambda_4 < 0$ for $\eta > 0$, and that all λ are positive for $\eta < 0$. From this, the boundary of the stable P state is given by

$$T = 1 \text{ for } \eta > 0, \tag{D·9}$$

$$\eta = \frac{4T(1-T)}{4T-1} \text{ for } \eta < 0. \tag{D·10}$$

(3) Mixed state M_4.

$$\Lambda_{11} = \Lambda_{22} = \Lambda_{33}, \Lambda_{12} = \Lambda_{13} = \Lambda_{23}, \Lambda_{14} = \Lambda_{24} = \Lambda_{34}.$$

$$|\Lambda - \lambda E| = \{\lambda - (\Lambda_{11} - \Lambda_{12})\}^2\{\lambda^2 - (\Lambda_{11} + \Lambda_{44} + 2\Lambda_{12})\lambda + 2\Lambda_{44}\Lambda_{12} + \Lambda_{11}\Lambda_{44} - 3\Lambda_{14}^2\} = 0. \tag{D·11}$$

(4) Mixed state M_5.

$$\Lambda_{11} = \Lambda_{22} = \Lambda_{33}, \Lambda_{13} = \Lambda_{23}, \Lambda_{14} = \Lambda_{24},$$

$$\begin{aligned}|\Lambda - \lambda E| &= \{\lambda - (\Lambda_{11} - \Lambda_{12})\}\{\lambda^3 - (2\Lambda_{11} + \Lambda_{44} + \Lambda_{12})\lambda^2 \\ &+(2\Lambda_{11}\Lambda_{44} + \Lambda_{11}\Lambda_{12} + \Lambda_{11}^2 + \Lambda_{44}\Lambda_{12} - 2\Lambda_{13}^2 - \Lambda_{34}^2 - 2\Lambda_{14}^2)\lambda \\ &-\Lambda_{11}\Lambda_{44}\Lambda_{12} - \Lambda_{11}^2\Lambda_{44} - 4\Lambda_{34}\Lambda_{13}\Lambda_{14} + 2\Lambda_{11}\Lambda_{14}^2 + 2\Lambda_{44}\Lambda_{13}^2 + \Lambda_{12}\Lambda_{34}^2 + \Lambda_{11}\Lambda_{34}^2\} = 0.\end{aligned}$$

(D·12)

At the boundary, some λ becomes 0. That is,

$$\Lambda_{11} = \Lambda_{12}, \tag{D·13}$$

or

$$-\Lambda_{11}\Lambda_{44}\Lambda_{12} - \Lambda_{11}^2\Lambda_{44} - 4\Lambda_{34}\Lambda_{13}\Lambda_{14} + 2\Lambda_{11}\Lambda_{14}^2 + 2\Lambda_{44}\Lambda_{13}^2 + \Lambda_{12}\Lambda_{34}^2 + \Lambda_{11}\Lambda_{34}^2 = 0. \tag{D·14}$$

(5) S_2 state.

$$\Lambda_{11} = \Lambda_{22} = \Lambda_{33}, \Lambda_{12} = \Lambda_{13}, \Lambda_{24} = \Lambda_{34},$$

$$\begin{aligned} |\Lambda - \lambda E| \quad = \quad & \{\lambda - (\Lambda_{11} - \Lambda_{23})\}\{\lambda^3 - (2\Lambda_{11} + \Lambda_{44} + \Lambda_{23})\lambda^2 \\ & +(2\Lambda_{11}\Lambda_{44} + \Lambda_{11}\Lambda_{23} + \Lambda_{44}\Lambda_{23} + \Lambda_{11}^2 - 2\Lambda_{24}^2 - 2\Lambda_{12}^2 - \Lambda_{14}^2)\lambda \\ & -\Lambda_{23}\Lambda_{11}\Lambda_{44} - \Lambda_{11}^2\Lambda_{44} - 4\Lambda_{12}\Lambda_{24}\Lambda_{14} + 2\Lambda_{11}\Lambda_{24}^2 + 2\Lambda_{44}\Lambda_{12}^2 + \Lambda_{14}^2\Lambda_{23} + \Lambda_{11}\Lambda_{14}^2\} = 0. \end{aligned} \tag{D·15}$$

At the boundary, some λ becomes 0. That is,

$$\Lambda_{11} = \Lambda_{23} \tag{D·16}$$

or

$$-\Lambda_{23}\Lambda_{11}\Lambda_{44} - \Lambda_{11}^2\Lambda_{44} - 4\Lambda_{12}\Lambda_{24}\Lambda_{14} + 2\Lambda_{11}\Lambda_{24}^2 + 2\Lambda_{44}\Lambda_{12}^2 + \Lambda_{14}^2\Lambda_{23} + \Lambda_{11}\Lambda_{14}^2 = 0. \tag{D·17}$$

References

1) F. C. Crick, and G. Mitchison, Nature **304**, 111 (1983).

2) J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

3) D. J. Amit, H. Gutfreund, and H. Somplinski, Phys. Rev. A **32**, 1007 (1985).

4) J. Hertz, A. Krogh, and R. G.Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, 1991).

5) D. E. Rumelhart, G. E. Hintn, and R. J. Williams, *Learning Internal Representations by Error Propagation, Parallel Distributed Processing, Exploration in the Microstructures of Cognition* (MIT Press, Cambridge, 1986) Vol. 1, Chap. 8, p. 318.

6) J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, Nature **304**, 158 (1983).

7) S. Wimbauer, N. Klemmer, and J. L. van Hemmen, Neural Networks **7**, 219 (1994).

8) M. C. D. Barrozo, and T. J. P. Penna, Int. J. Mod. Phys. C **5**, 503 (1994).

9) S. Wimbauer, and J. L. van Hemmen, in *Proceedings on Analysis of Dynamical and Cognitive Systems, Advanced Course* (Springer-Verlag, London, 1995) p. 121.

10) J. A. Horas, and E. A. Bea, Int. J. Neur. Syst. **12**, 109 (2002).

11) K. Nokura, Phys. Rev. E **54**, 5571 (1996).

12) K. Nokura: J. Phys. A: Math. Gen. **31**, 7447 (1998).