PAPER

Associative Memory Model with Forgetting Process Using Nonmonotonic Neurons

Kazushi MIMURA†, Masato OKADA††, *and* Koji KURATA†, *Members*

SUMMARY An associative memory model with a forgetting process *a la* Mézard et al. is investigated for a piecewise nonmonotonic output function by the SCSNA proposed by Shiino and Fukai. Similar to the formal monotonic two-state model analyzed by Mézard et al., the discussed nonmonotonic model is also free from a catastrophic deterioration of memory due to overloading. We theoretically obtain a relationship between the storage capacity and the forgetting rate, and find that there is an optimal value of forgetting rate, at which the storage capacity is maximized for the given nonmonotonicity. The maximal storage capacity and capacity ratio (a ratio of the storage capacity for the conventional correlation learning rule to the maximal storage capacity) increase with nonmonotonicity, whereas the optimal forgetting rate decreases with nonmonotonicity.
key words: *associative memory model, nonmonotonic neuron, forgetting process, storage capacity, SCSNA*

1. Introduction

It is well known that the maximal number of memory patterns for the Hopfield associative memory model [1] is $0.14N$ [2], where N is the number of neurons. A ratio of the maximal number of memory patterns to the neuron number is called storage capacity. If the number of memory patterns exceeds $0.14N$, all previously stored memory patterns become unstable, namely there is a catastrophic deterioration of memory. In order to avoid this, one can introduce a forgetting mechanism into the learning process,

$$J_{ij}^t = \eta J_{ij}^{t-1} + \frac{1}{N}\xi_i^t \xi_j^t, \qquad (1)$$

where J_{ij} is the synaptic weight from the j-th neuron to the i-th neuron, ξ_i^t is the i-th component of the t-th memory pattern, and J_{ij}^t is the synaptic weight after embedding the t-th memory pattern. The introduced forgetting mechanism can be obtained by denoting $\eta < 1$, whereas Eq. (1) reproduces the conventional correlation type learning rule in the case of $\eta = 1$, which corresponds to the Hebbian rule. It could be expected that only the recently stored p_C memory patterns are stable, whereas the rest memory patterns are unstable since $\eta < 1$. As discussed in the next section, there is an optimal value of forgetting rate, at which the p_C is maximized. Actually, this model was analyzed by Mézard et al. using the replica theory [3]. They discussed the relationship between p_C and η, and obtained the optimal value of forgetting rate and the maximal p_C. Okada et al. [4] also analyzed the same model by the statistical neurodyanmics [5]. They also succeeded in obtaining a similar relationship between p_C and η, and the optimal value of forgetting rate and the maximal p_C.

It has been shown that the storage capacity of an associative memory model can be improved markedly by replacing the usual sigmoid neurons with nonmonotonic ones [6]. Yoshizawa et al. [7] showed that the storage capacity of an associative memory model with optimal nonmonotonicity is 0.4, which is approximately three times as large as that of the Hopfield model [2]. However, when the number of memory patterns exceeds the storage capacity, all previously stored memory patterns become unstable even in the nonmonotonic model. In the present paper, we analyze a nonmonotonic model with a forgetting process by using the self consistent signal to noise analysis (SCSNA) proposed by Shiino and Fukai [8], [9]. A piecewise linear model of the nonmonotonic neuron is adopted [10]. We discuss the relationship between p_C and η, and the maximal p_C at the optimal forgetting rate.

2. Model

We begin by formulating a recurrent neural network with N analogue neurons that have an output function $F(\cdot)$. The network model is studied with an infinite number of neurons, i.e., $N \to \infty$. The network dynamics are written with the internal potential u variables as

$$\frac{d}{dt}x_i = -x_i + F(u_i),$$
$$u_i = \sum_{j \neq i}^{N} J_{ij} x_j, \qquad (2)$$

where x_i denotes an output of the i-th neuron. The internal potential u variables are assumed to be connected with each other through the synaptic weights J_{ij} of a form

Manuscript received January 16, 1998.
Manuscript revised May 13, 1998.
†The authors are with the Department of Systems and Human Science, Graduate School of Engineering Science, Osaka University, Toyonaka-shi, 560–8531 Japan.
††The author is with Kawato Dynamic Brain Project, ERATO, Japan Science and Technology Corporation (JST), Kyoto-fu, 619–0288 Japan.

$$J_{ij} = \frac{1}{N}\sum_{\mu=0}^{\infty}\eta^{\mu}\xi_i^{\mu}\xi_j^{\mu}, \tag{3}$$

where ν is the serial number of the memory pattern, so that $\nu = 0$ corresponds to the latest stored pattern, and the larger serial number μ means that the pattern μ was stored earlier. Equation (3) is equivalent to the infinite operations result of Eq. (1). Each element ξ_i^{μ} of the μ-th memory pattern is an independent random variable that takes a value of 1 or -1 with a probability

$$\mathrm{Prob}[\xi^{\mu} = \pm 1] = \frac{1}{2}. \tag{4}$$

Let us discuss how p_c and η depend on the number of neurons. The equilibrium of Eq. (2) is obtained by denoting $du_i/dt = 0$. Assuming that the ν-th memory pattern is rigorously retrieved, i.e., $x_i = \xi_i^{\nu}$, the internal potential of the i-th neuron u_i in the equilibrium state is

$$u_i = \sum_{j\neq i}^{N} J_{ij}\xi_j^{\nu} \tag{5}$$

$$= \eta^{\nu}\xi_i^{\nu} + \frac{1}{N}\sum_{\mu=0,\mu\neq\nu}^{\infty}\sum_{j\neq i}^{N}\eta^{\mu}\xi_i^{\mu}\xi_j^{\mu}\xi_j^{\nu}, \tag{6}$$

where the first term on the right-hand side (RHS) is the retrievable signal, whereas the second term is the cross-talk noise which prevents $\boldsymbol{\xi}^{\nu}$ from being retrieved. The mean of the cross-talk noise is 0 and its variance is

$$\frac{1}{N}\sum_{\mu=0,\mu\neq\nu}^{\infty}\sum_{j\neq i}^{N}(\eta^{2\mu}) = \frac{1}{N(1-\eta^2)}. \tag{7}$$

We can estimate that the variance should be $O(1)$ with respect to the number of neurons N from a finding of the conventional correlation learning rule in the Hopfield model [1], [2]. This estimation gives $\eta \sim 1 - O(1/N)$ and $p_C \sim O(N)$. According to this result, we rewrite η and ν as

$$\eta = \exp\left(-\frac{\varepsilon^2}{2N}\right) \tag{8}$$

$$\alpha = \frac{\nu}{N} \tag{9}$$

where we define ε as the forgetting rate and α as the loading rate. The loading rate is a ratio of the sequential number of memory pattern to the number of neurons. Storage capacity α_C is also defined as $\alpha_C = p_C/N$ using the maximal number of retrievable patterns. Using the forgetting rate ε and the loading rate α, Eqs. (6) and (7) are rewritten as

$$u_i = \exp\left(-\frac{\alpha\varepsilon^2}{2}\right)\xi_i^{\alpha} + \frac{1}{\varepsilon^2}z_i, \tag{10}$$

where we replace ξ_i^{ν} with ξ_i^{α} and z_i obeys the Gaussian distribution with mean 0 and variance 1. The S/N ration between the signal and the cross-talk noise in Eq. (10) is $\varepsilon\exp(-\alpha\varepsilon^2/2)$, which has an maximal value for $\varepsilon > 0$. This fact implies that the storage capacity is determined by not only the amount of the signal term but also one of the cross-talk noise, i.e., competition between them, and that there is an optimal value of forgetting rate, at which the storage capacity α_C is maximized.

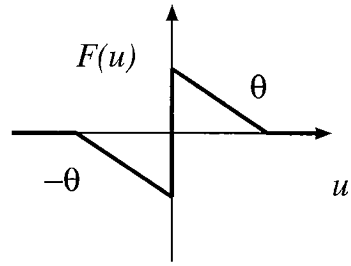


Fig. 1 Piecewise linear nonmonotonic output function.

We discuss the following generalized Hebbian learning rule in the present paper including the previously discussed forgetting case,

$$J_{ij} = \frac{1}{N}\sum_{\mu=0}^{\infty}\Lambda\left(\frac{\mu}{N}\right)\xi_i^{\mu}\xi_j^{\mu}. \tag{11}$$

Denoting

$$\Lambda(s) = \exp\left(-\frac{\varepsilon^2}{2}s\right) \tag{12}$$

and

$$\Lambda(s) = \begin{cases} 1 & 0 \leqq s \leqq \alpha \\ 0 & s > \alpha \end{cases} \tag{13}$$

provides the forgetting case and the conventional correlation learning rule case, respectively. In this paper, we use the following odd function $F(u)$ in Fig. 1 as an output function

$$F(u) = \begin{cases} -\frac{1}{\theta}u - 1, & -\theta < u < 0, \\ -\frac{1}{\theta}u + 1, & 0 < u < \theta, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

In this paper, we use $1/\theta$ as a parameter representing a degree of nonmonotonicity. In the case of $1/\theta \to 0$, the output function $F(\cdot)$ in Eq. (14) converges on the sign function $\mathrm{sgn}(\cdot)$.

The overlap between the μ-th memory pattern $\boldsymbol{\xi}^{\mu}$ and the equilibrium network state $\boldsymbol{x}$ is now defined as

$$m_{\mu} = \frac{1}{N}\sum_{j=1}^{N}\xi_j^{\mu}x_j. \tag{15}$$

In the case the output function $F(\cdot)$ is set to the sign function sgn(·), the overlap m_μ represents the direction cosine between the memory pattern $\boldsymbol{\xi}^\mu$ and the network state $\boldsymbol{x}$. We focus on the equilibrium in which only the $\nu(=\alpha N)$-th pattern is retrieved, i.e., $m_\alpha \equiv m_\nu = O(1)$ and $m_\mu = O(1/\sqrt{N})$, where $\mu \neq \nu$.

3. Results

3.1 SCSNA

The internal potential u_i of each neuron in the equilibrium state is represented by the weighted sum of outputs from other neurons. The basis of the SCSNA is in the systematic splitting of the internal potential into a signal and a cross-talk noise. Moreover, the cross-talk noise part consists of two elements. One is an effective self-coupling term Γ, which comes from statistical correlations caused by the recurrent connections, the other obeys the Gaussian distribution with a mean 0 and a variance σ^2. The following results are obtained for any odd output function $F(\cdot)$

$$m_\alpha = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) Y(\alpha, z), \tag{16}$$

$$q = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) Y(\alpha, z)^2, \tag{17}$$

$$U = \frac{1}{\sigma} \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) zY(\alpha, z), \tag{18}$$

$$Y(\alpha, z) = F\left(\Lambda(\alpha) m_\alpha + \Gamma Y(\alpha, z) + \sigma z\right), \tag{19}$$

$$\sigma^2 = q \int_0^{\infty} ds \frac{\Lambda(s)^2}{(1 - \Lambda(s)U)^2}, \tag{20}$$

$$\Gamma = \int_0^{\infty} ds \frac{\Lambda(s)^2 U}{1 - \Lambda(s)U}, \tag{21}$$

where $Y(\alpha, z)$ is an effective output function of each neuron obtained by solving Eq. (19), m_α is the overlap for the target pattern $\boldsymbol{\xi}^\nu$ ($\nu = \alpha N$) defined in Eq. (15), and q is the so-called Edwards-Anderson order parameter. The susceptibility U measures the sensitivity of neuron output to external input, and σ^2 is the noise variance that obeys the Gaussian distribution. In the forgetting case in Eq. (12) [3],

$$\sigma^2 = \frac{2q}{\varepsilon^2} \left[\frac{1}{U^2} \log_e(1 - U) + \frac{1}{U(1 - U)}\right] \tag{22}$$

$$\Gamma = -\frac{2}{\varepsilon^2} \left[\frac{1}{U} \log_e(1 - U) + 1\right] \tag{23}$$

whereas in the conventional correlation learning rule in Eq. (13) [8]–[10],

$$\sigma^2 = \frac{\alpha q}{(1 - U)^2} \tag{24}$$

$$\Gamma = \frac{\alpha U}{1 - U}. \tag{25}$$

According to the SCSNA, the macroscopic description of any microscopic state can be represented by the three order parameters of m_α, q, and U. The storage capacity can be calculated by solving Eqs. (16)–(21) self-consistently as follows. The order parameter equations are solved numerically. There is a transition at $\alpha = \alpha_C$. When the loading rate α is below α_C, Eqs. (16)–(21) have a nontrivial solution with $m_\alpha \neq 0$. On the other hand, if α is above α_C, only the trivial solution with $m_\alpha = 0$ exists. The critical value α_C is the storage capacity. Derivation of the above-mentioned order-parameter equations of Eqs. (16) through (21) is given in Appendix.

3.2 Conventional Correlation Learning

Before analyzing the forgetting model, we show the storage capacity of the model with the conventional correlation learning rule in Eq. (13) [10] for comparison. Figure 2 illustrates the relationship of the storage capacity α_C to the nonmonotonicity $1/\theta$ for the conventional correlation learning rule. As previously mentioned, the Hopfield model corresponds to $1/\theta = 0$. $1/\theta = 0.5$ gives the highest nonmonotonicity because the storage capacity obtained by the SCSNA in Eq. (13) tends to be larger than the actual values obtained by computer simulation of $1/\theta > 0.5$ [10]. Since the susceptibility U in Eq. (18) is always positive in the Hopfield model ($1/\theta = 0$), the variance of the cross-talk noise σ^2 in Eqs. (20) or (24) is always larger than αq. In the nonmonotonic model, however, U may be negative in the retrieval phase. Therefore, the variance of the cross-talk noise σ^2 is smaller than the variance in the Hopfield model. The storage capacity of the nonmonotonic model monotonically increases with the nonmonotonicity $1/\theta$ in $0 \leq 1/\theta \leq 0.5$.

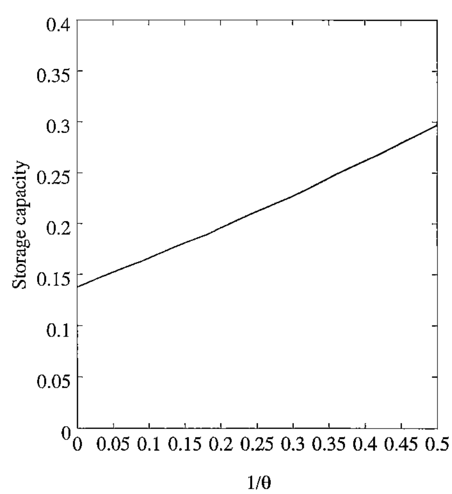


Fig. 2 Storage capacity α_C and nonmonotonicity $1/\theta$ for the conventional correlation learning rule.

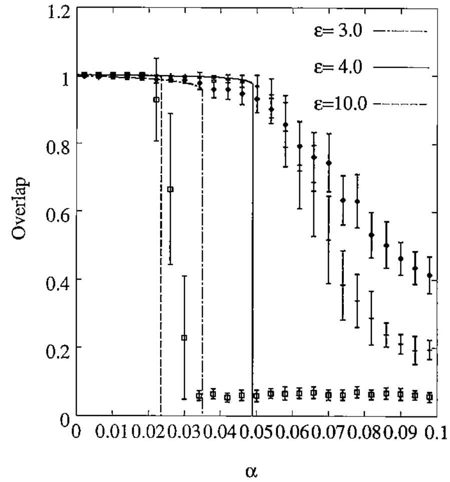


Fig. 3 Theoretically obtained overlap m_α and loading rate α for forgetting rate values $\varepsilon = 3.0, 4.0$ and 6.0 on the monotonic two-state network $1/\theta = 0$. Experimental results of overlap are also shown.

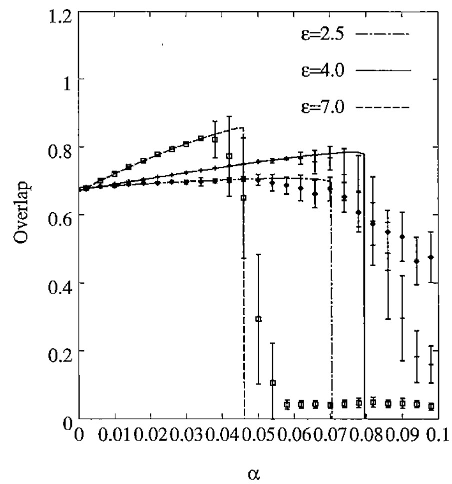


Fig. 4 Theoretically obtained overlap m_α and loading rate α for forgetting rate values $\varepsilon = 2.5, 4.0$ and 7.0 on the nonmonotonic network $1/\theta = 0.5$. Experimental results of overlap are also shown.

3.3 Forgetting Process

Mézard et al. proposed the forgetting process for the learning rule of Eqs. (11) and (12), and analyzed their model of two-state (± 1) neurons by the replica theory [3]. The replica theory results coincide with previously discussed analytical results of the SCSNA. Similarly, the result of the SCSNA from Eqs. (16) through (21) also coincides with the result of Mézard et al., if we set $F(\cdot) = \mathrm{sgn}(\cdot)$. Figure 3 shows the overlap m_α of the retrieved $\nu = \alpha N$-th pattern of the monotonic two-state model, i.e., $F(\cdot) = \mathrm{sgn}(\cdot)$, and this implies that there is an optimal forgetting rate $\varepsilon_{\mathrm{opt}}$ because the storage capacity of $\varepsilon = 4.0$ has the largest value. The dash-dotted line in Fig. 5 shows the storage capacity α_C of the monotonic two-state network as a function of the forgetting rate ε. As implied in Fig. 3, the storage capacity α_C has a maximal value of 0.049 at $\varepsilon = 4.1$, and this is exactly equivalent to the one obtained by Mézard et al. [3].

Figure 4 shows theoretical results of the overlap m_α of the retrieved $\nu = \alpha N$-th pattern in the highest nonmonotonic case, i.e., $1/\theta = 0.5$, and also implies that there is an optimal forgetting rate $\varepsilon_{\mathrm{opt}}$ similar to the monotonic case, even if the network dynamics are nonmonotonic. Because generally $F(u) \leq 1$, the overlap is less than 1 in the nonmonotonic case. To confirm these theoretical results, numerical simulations were carried out as shown in Figs. 3 and 4. We employed a discrete-time version of Eq. (2), i.e.,

$$x_i(t + \Delta t) = (1 - \Delta t)x_i(t) + \Delta t F(u_i(t)),$$
$$u_i(t) = \sum_{j \neq i}^{N} J_{ij} x_j(t), \qquad (26)$$

with $\Delta t = 0.1$. No numerical improvement was seen in further decreasing Δt. Date points representing the overlap m_α are obtained by computer simulation with $N = 500$ for 50 trials in the both figures. Error bars indicate standard deviations. In comparing Figs. 3 and 4, one may find that the overlap m_α depends on the loading rate α in different way. The overlap m_α decreases as the loading rate α increases in the monotonic two-state network and increases in the nonmonotonic case. The reason is as follows: Since the loading rate corresponds to the sequential number of the memory pattern, the larger loading rate means that the corresponding memory pattern is stored earlier. The signal term of the first term in the RHS of Eq. (19) is scaled as $\Lambda(\alpha) = \exp(-\varepsilon^2\alpha/2)$. This finding implies that both the larger ε and α result in a smaller coefficient of the signal term, and that the nonmonotonicity is effectively decreased as ε and/or α is increased. Thus the network with the larger values of ε and α has effectively weaker nonmonotonicity and a larger overlap value. If we compare the results of $\varepsilon = 2.5$, 4.0 and 7.0, the overlap m_α increases the most with the largest value of ε. Figure 5 shows the storage capacity α_C for various values of the nonmonotonicity $1/\theta$ including the monotonic two-state network as a function of the forgetting rate ε. As in the monotonic case, the storage capacity α_C has the maximal value $\alpha_C(\varepsilon_{\mathrm{opt}})$ at a certain value of $\varepsilon_{\mathrm{opt}}$ depending on the nonmonotonicity.

Figures 6 and 7 show the maximal storage capacity $\alpha_C(\varepsilon_{\mathrm{opt}})$ and the optimal forgetting rate $\varepsilon_{\mathrm{opt}}$ for various nonmonotonicity values, i.e., $0 \leq 1/\theta \leq 0.5$. As previously mentioned, $1/\theta = 0.5$ is the maximal amount of the nonmonotonicity where both analytical and computer simulation results coincide [10]. As with the conventional correlation learning rule shown in Fig. 2, the maximal storage capacity $\alpha_C(\varepsilon_{\mathrm{opt}})$ increases with the

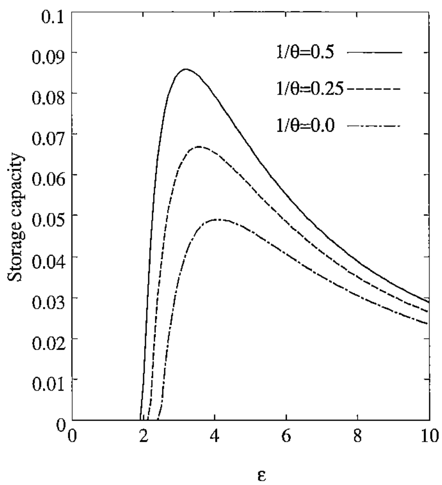


Fig. 5 Storage capacity α_C and forgetting rate ε for various values of nonmonotonicity $1/\theta = 0.0, 0.25$ and 0.5.

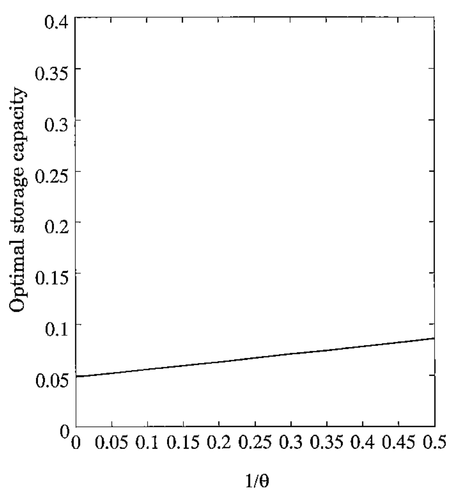


Fig. 6 Maximal storage capacity $\alpha_C(\varepsilon_{\rm opt})$ for the forgetting process and nonmonotonicity $1/\theta$.

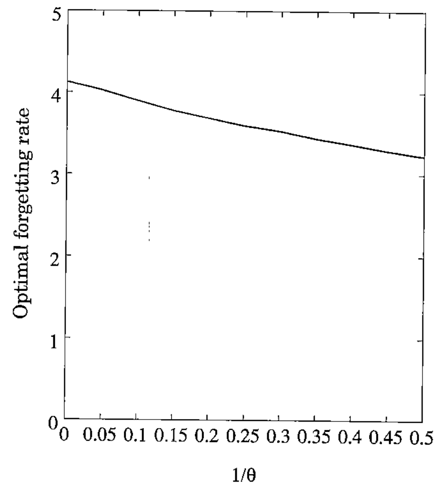


Fig. 7 Optimal forgetting rate $\varepsilon_{\rm opt}$ and nonmonotonicity $1/\theta$.

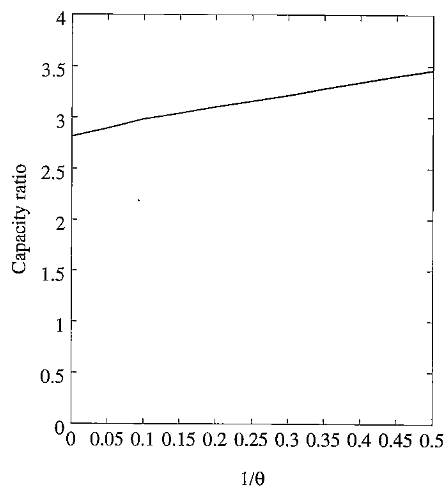


Fig. 8 Capacity ratio $\alpha_C/\alpha_C(\varepsilon_{\rm opt})$ and nonmonotonicity $1/\theta$.

nonmonotonicity $1/\theta$. On the other hand, the optimal forgetting rate $\varepsilon_{\rm opt}$ decreases with the nonmonotonicity $1/\theta$. This indicates that the coefficient of the signal part $\Lambda(\alpha)$ of Eq. (19) at $\alpha = \alpha_C(\varepsilon_{\rm opt})$ tends to be constant with respect to the variation of the nonmonotonicity.

Mézard et al. discussed a capacity ratio of the storage capacity α_C of the conventional model in Eq. (13) with respect to the maximal storage capacity $\alpha_C(\varepsilon_{\rm opt})$ on the forgetting process ($\alpha_C/\alpha_C(\varepsilon_{\rm opt})$). They showed $\alpha_C/\alpha_C(\varepsilon_{\rm opt}) \approx 2.82$ for the monotonic two-state network [3]. Akaho [11] discussed the same model concerning the absolute capacity [5], [12], which is an upper limit on the number of memory patterns to guarantee errorless retrieval. The absolute capacity of the conventional correlation learning rule is $N/(2\log_e N)$ [5], [12], while the maximal absolute capacity for the forgetting process is $N/(2e\log_e N)$ at $\eta = 1 - (e\log_e N)/N$. Thus the capacity ratio concerning the absolute capacity is $e \approx 2.72$. This value approximately coincides with one obtained by Mézard et al. Figure 8 shows the capacity ratio $\alpha_C/\alpha_C(\varepsilon_{\rm opt})$ as a function of the nonmonotonicity $1/\theta$, and demonstrates that the capacity ratio increases with the nonmonotonicity. This means that the storage capacity α_C of the forgetting case is not so enhanced by introducing the nonmonotonicity as one of the conventional model. The reason is that the *effective* nonmonotonicity at $\alpha = \alpha_C(\varepsilon_{\rm opt})$, which can be considered to be $\Lambda(\alpha_C(\varepsilon_{\rm opt}))/\theta$, is relatively smaller than the *proper* nonmonotonicity $1/\theta$.

4. Conclusion

In this paper, we have investigated the properties of the associative memory model with the forgetting process for the piecewise linear nonmonotonicity by the SCSNA. When the nonmonotonicity disappears, the obtained SCSNA order-parameter equation completely concurs with the analysis of the monotonic two-state network by Mézard et al. [3]. The discussed model is free from the catastrophic deterioration of memory due to overloading. We have shown the relationship between the storage capacity and the forgetting rate, and obtained the optimal forgetting rate, at which is the storage capacity is maximized. The maximal storage capacity and the capacity ratio increase with the nonmonotonicity, whereas the optimal forgetting rate decreases with it.

References

[1] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc. National Academy of Sciences, vol.79, pp.2554–2558, 1982.

[2] D.J. Amit, H. Gutfreund, and H. Sompolinsky, "Storing infinite numbers of patterns in a spin-glass model of neural networks," Physical Review Letters, vol.55, pp.1530–1533, 1985.

[3] M. Mézard, J.P. Nadal, and G. Toulouse, "Solvable models of working memories," J. Physique, vol.47, pp.1457–1462, 1986.

[4] M. Okada, K. Mimura, and K. Kurata, "Associative memory with forgetting process — Analysis by statistical neurodynamics," IEICE Trans., vol.J77-D-II, pp.1178–1180, 1994.

[5] S. Amari and K. Maginu, "Statistical neurodynamics of associative memory," Neural Networks, vol.1, pp.63–73, 1988.

[6] M. Morita, S. Yoshizawa, and K. Nakano, "Analysis and improvement of the dynamics of autocorrelation associative memory," IEICE Trans., vol.J73-D-II, pp.232–242, 1990.

[7] S. Yoshizawa, M. Morita, and S. Amari, "Capacity of associative memory using a nonmonotonic neuron model," Neural Networks, vol.6, pp.167–176, 1993.

[8] M. Shiino and T. Fukai, "Self-consistent signal-to-noise analysis and its application to analogue neural networks with asymmetric connections," Journal of Physics A: Mathematical and General, vol.25, pp.L375–L381, 1992.

[9] M. Shiino and T. Fukai, "Self-consistent signal-to-noise analysis of the statistical behavior of analog neural networks and enhancement of the storage capacity," Physical Review E, vol.48, pp.867–897, 1993.

[10] T. Fukai, J. Kim, and M. Shiino, "Retrieval properties of analog neural networks and the nonmonotonicity of transfer functions," Neural Networks, vol.8, pp.391–404, 1995.

[11] S. Akaho, "Optimal decay rate of connection weights in covariance learning," Proc. 1994 Annual Conference of Japanese Neural Networks Society, pp.129–130, 1992.

[12] R.J. McEliece, E.C. Posner, E.R. Rodemich, and S.S. Venkatesh, "The capacity of the Hopfield associative memory," IEEE Trans. Inf. Theory, vol.IT-33, pp.461–482, 1987.

Appendix: Derivation of the SCSNA: The Forgetting Process

Expressing the internal potential u_i in the equilibrium state by the overlaps (15), we obtain

$$u_i = \sum_{j \neq i}^{N} J_{ij} x_j = \sum_{\mu=0}^{\infty} \Lambda\left(\frac{\mu}{N}\right) \xi_i^\mu m_\mu - \int_0^\infty ds \Lambda(s) x_i. \qquad \text{(A· 1)}$$

The output x_i can be formally expressed as

$$x_i = F\left(\sum_{\mu=0}^{\infty} \Lambda\left(\frac{\mu}{N}\right) \xi_i^\mu m_\mu - \int_0^\infty ds \Lambda(s) x_i\right) = \tilde{F}\left(\sum_{\mu=0}^{\infty} \Lambda\left(\frac{\mu}{N}\right) \xi_i^\mu m_\mu\right),$$

where the function $\tilde{F}(\cdot)$ will be determined later. Here, we discuss when the ν-th pattern $\boldsymbol{\xi}^\nu$ is retrieved. The residual overlap $m_\mu = O(1/\sqrt{N})$, $(\mu \neq \nu)$ is obtained by using the Taylor expansion

$$m_\mu = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \tilde{F}\left(\sum_{\rho=0}^{\infty} \Lambda\left(\frac{\rho}{N}\right) \xi_i^\rho m_\rho\right) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu x_i^{(\mu)} + \Lambda\left(\frac{\mu}{N}\right) U m_\mu, = \frac{1}{N(1-\Lambda\left(\frac{\mu}{N}\right)U)} \sum_{i=1}^{N} \xi_i^\mu x_i^{(\mu)} \qquad \text{(A· 2)}$$

where

$$x_i^{(\mu)} = \tilde{F}\left(\sum_{\rho \neq \mu}^{\infty} \Lambda\left(\frac{\rho}{N}\right) \xi_i^\rho m_\rho\right),$$

$$x_i^{\prime(\mu)} = \tilde{F}^\prime\left(\sum_{\rho \neq \mu}^{\infty} \Lambda\left(\frac{\rho}{N}\right) \xi_i^\rho m_\rho\right),$$

$$U = \frac{1}{N} \sum_{i=1}^{N} x_i^{\prime(\mu)}.$$

The sequential number of the retrieved pattern $\boldsymbol{\xi}^\nu$, i.e. ν, should be scaled with the number of neurons N, that is, $\nu = \alpha N$. We call this α the loading rate. Substituting Eq. (A· 2) into Eq. (A· 1), we get

$$u_i = \Lambda(\alpha) \xi_i^\alpha m_\alpha + \int_0^\infty ds \frac{\Lambda(s)^2 U}{1-\Lambda(s)U} x_i + \bar{z},$$

where we replaced the signal part of the internal potential:

$$\xi_i^\nu \to \xi_i^\alpha, \quad m_\nu \to m_\alpha,$$

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

$$\bar{z} = \frac{1}{N} \sum_{\mu \neq \nu}^{\infty} \sum_{j \neq i}^{N} \frac{\Lambda(\frac{\mu}{N})}{1 - \Lambda(\frac{\mu}{N})} \xi_i^{\mu} \xi_j^{\mu} x_j^{(\mu)}.$$

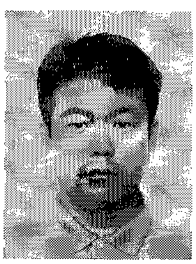
Note that $\bar{z}$ is a summation of uncorrelated random variables, with $< \bar{z} >= 0$ and $< \bar{z} >= \sigma^2$. Thus,

$$\sigma^2 = q \int_0^{\infty} ds \frac{\Lambda(s)^2}{(1 - \Lambda(s))^2},$$

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

If we replace $u_i \rightarrow u$, $\xi_i^{\alpha} m_{\alpha} \rightarrow m_{\alpha}$, and $\tilde{F}(u) \rightarrow Y$, and denote $z = \bar{z}/\sigma$, and since we have discussed the odd function $F(\cdot)$, Eqs. (16)–(21) are obtained.

Kazushi Mimura received the B.Eng. and M.Eng. degrees from Osaka University, Japan, in 1992 and 1994, respectively. He has been a student of Graduate School of Engineering Science, Osaka University since 1994.



Masato Okada received the B.Sci. degree in physics from Osaka City University, Japan, in 1985. He received the M.Sci. degree in physics from Osaka University, Japan, in 1987. From 1987 to 1989 he joined Mitsubishi Electric Corporation. From 1989 to 1991 he was a student of Graduate School of Engineering Science, Osaka University. From 1991 to 1996 he was a Research Associate at Osaka University. He received the Ph.D. degree in science from Osaka University in 1997. He has been a researcher in Kawato dynamics brain project supported by Japan Science and Technology corporation since 1996. His research interests include computational aspects of neural networks, especially those of vision and memory.

Koji Kurata received the B.S., M.S., and Ph.D. degrees in mathematical engineering from the University of Tokyo, Japan, in 1981, 1983, and 1990 respectively. From 1984 to 1990 he was a Research Associate at the University of Tokyo. He has been an Assistant Professor at Osaka University since 1990. His research interests include computational aspects of neural networks, especially those of topological mapping organization. He is also interested in pattern formations in homogeneous fields.