数理解析研究所講究録
第 1742 巻 2011 年 145-157

Complex-Sigmoid Function for Complex-Valued Neural Networks

東京大学・生産技術研究所　田中　剛平 (Gouhei Tanaka)
Institute of Industrial Science, The University of Tokyo

Abstract

A complex-valued neural network is a class of artificial neural network for dealing with complex-valued and multivalued data. It consists of complex-valued neurons, complex-valued weights, and complex activation functions. Since complex number representation and arithmetic are appropriate for processing complex-valued information like wave phenomena as well as multivalued information like digital images, complex-valued neural networks are thought to have a potential for wide applications. Yet, nonlinear dynamics of complex-valued neural networks has not been fully understood. In this report, we first overview the complex-valued Hopfield network for multistate associative memories. Second, we reveal some properties of the complex-sigmoid function and discuss the effect of its nonlinearity parameter influencing the network capability.

1 Introduction

Artificial neural network is one of the major frameworks in computational intelligence and machine learning [1]. It is often difficult to strictly solve a problem in the real world, even if the problem can be formulated mathematically. In such a case, an effort would be made to numerically search a better solution. Computational intelligence aims to explore and develop effective algorithms to find a good solution in a feasible computation time. Even when there is a method to find the best solution, it does not make sense practically if the method takes a lot of time and cost. To explore a reasonable algorithm, neural networks have been extensively studied with various theoretical and heuristic methods.

In conventional binary and real-valued neural networks, the signum (Heaviside step) function and the sigmoid function have been typically used as an activation function. The signum function returns one of the two states (corresponding to neuronal firing and non-firing states) depending on the input and the threshold parameter. This two-state neuron is suited to address binary information. The Hopfield neural networks with recurrent connections have been widely applied to pattern recognition and combinatorial optimization formulated by using binary variables [2]. The sigmoid function, which is a continuous version of the signum function, enables to derive the backpropagation learning algorithm [3]. Also in associative memories, the continuous nonlinearity of the sigmoid function can be effective to prevent a neuronal state from being trapped by a local spurious pattern as often seen in the network with the signum function.

When two-state neural networks are applied to some problems related to multivalued information, we need to preliminary transform the multivalued information into binary information and vice versa after information processing. However, it is often not straightforward but heuristic and thereby can be a cause of unnecessary errors. Therefore, many attempts have been made to develop neural network models suited for multivalued information processing. Multivalued information can be directly represented by multistate neurons which require fewer neurons and fewer interconnections than two-state neurons. It is desirable in VLSI implementation and optical implementation of the neural network.

A multistate neuron can be achieved by a multilevel activation function operating in a real space. In fact, multivalued neural networks and generalized Hopfield networks have been constructed for multistate associative memory models by using multilevel functions [4, 5, 6, 7, 8, 9]. They are applicable to gray-level image processing, but limited to the case where the number of gray levels is not so large.

From the viewpoint of topology, it is reasonable to code multiple equivalent states by the equiangularly divided points on the unit circle. This notion was first proposed in the 1970s [10] and developed in the studies on complex-valued neural networks [11, 12, 13, 14, 15, 16, 17]. The complex activation function transforms a complex-valued input on the complex domain into a complex-valued output on the unit circle. A generalization of the Hopfield network has been proposed as a complex-valued Hopfield network (CHN) for a multistate associative memory by using the complex-signum (CSIGN) activation function [18]. After that, the CHN has been modified to enhance its performance. The modifications include the introduction of the penalty term in the energy function [13], the learning rule which guarantees the strict local stability of the stored patterns [19], the connection scheme through the bifurcation parameters [20], the generalized projection learning rule which orthogonalizes the stored patterns [21], the complex-sigmoid (CSIGM) function which is an alternative to the CSIGN function [22].

In the previous study [22], we have shown that the CSIGN function with discrete nonlinearity can be generalized into the CSIGM function with continuous nonlinearity. This activation function highly improves the performance of the CHN in multistate associative memory tasks, resulting in the successful application to gray-level image restoration. However, the property of the CSIGM function has not yet been fully known. Therefore, we clarify some properties of the CSIGM function and the effect of its nonlinearity parameter.

In Sec. 2 we overview the CHN. In Sec. 3 we show some mathematical properties of the CSIGM function, including the effect of the nonlinearity parameter. Finally the results are summarized in Sec. 4.

2 Complex-valued Hopfield network

We briefly overview the complex-valued Hopfield networks (CHNs) for multistate associative memory [18, 22], which has a recurrent network structure with all-to-all interconnections. Let us consider a network composed of N complex-valued neurons. The state of the nth neuron is denoted by a complex variable z_n with $|z_n| = 1$ for $n = 1, \ldots, N$. The network state is represented by a vector $\mathbf{z} = (z_1, \ldots, z_N)$. The complex-valued weight coefficients are represented by w_{nj} for $1 \leq n \leq N$ and $1 \leq j \leq N$. The K-valued patterns to be stored are given by a set of P complex vectors $\mathbf{s}^{(p)} = (s_1^{(p)}, \ldots, s_N^{(p)})$ for $p = 1, \ldots, P$, where $s_n^{(p)} \in \{e^{ik\theta_K} \mid k = 0, \ldots, K-1\}$ and $\theta_K = 2\pi/K$. There are several choices for the learning rule and the update scheme as introduced below.

(a) Learning rule For associative memory models and image restoration tasks, the weight parameters are determined from a given pattern matrix $S = (\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(P)})$. The generalized Hebbian rule [11, 18], which is a generalization of the original Hebbian rule for binary networks [2], can be used for multistate neural networks as follows:

$$W = \frac{1}{N} S S^*, \tag{1}$$

where $W = (w_{nj})$ is the weight matrix and S^* is the complex conjugate transpose of S. The weight matrix W is Hermitian.

When the correlations among stored patterns are relatively large, it can lead to a recall of an incorrect undesirable pattern. The correlations can be reduced by orthogonalizing the stored patterns. This idea yields the generalized projection rule [14, 21, 22] as follows:

$$W = \frac{1}{N} S S^{+}, \tag{2}$$

where $S^{+} = (S^{*}S)^{-1}S^{*}$ is the Moore-Penrose pseudoinverse matrix of S.The diagonal entries of W are often removed to exclude the self-connection.

There is another learning rule which guarantees the strict local stability of the stored patterns [19]. This method requires to solve linear inequalities simultaneously by a linear programming procedure.

(b) Update scheme Each complex-valued neuron receives a sum of weighted inputs and then transforms it into an output by its complex activation function F. The nth neuron is updated as follows:

$$z'_n = F\left(e^{i\theta_K/2} \sum_{j=1}^{N} w_{nj} z_j \right), \tag{3}$$

where the term $e^{i\theta_K/2}$ is a factor to locate the resulting state in an angular center of the sector to which the state belongs. This update procedure is iterated asynchronously for a randomly chosen neuron until the network state converges or the maximum number of iterations is reached. The resulting states are transformed into multivalued information by the following rule: If $k\theta \leq \arg(z_n) < (k+1)\theta$, then return k.

The CSIGN function [11, 18] with discrete nonlinearity has been typically used as an activation function, which is a generalization of the signum function. An alternative to the CSIGN function is the CSIGM function [22] with continuous nonlinearity, which is a generalization of the sigmoid function. The CSIGM function includes the CSIGN one as a special case. In the next section, some mathematical properties of the CSIGM function are studied.

(c) Energy function The CHN has a well-defined energy function [18]:

$$E(\mathbf{z}) = -\frac{1}{2} \mathbf{z}^{*} W \mathbf{z}, \tag{4}$$

which is real-valued if the weight matrix W is Hermitian.

We first consider the CHN with the CSIGN activation function. We assume that the network state changes from $\mathbf{z}$ to $\mathbf{z}'$ only due to an update of the nth neuronal state from z_n to z'_n. Then, it is proved that the change of the energy function is not positive, i.e. $\Delta E = E(\mathbf{z}') - E(\mathbf{z}) \leq 0$ [18]. This guarantees that the network state converges to a certain state at a local minimum of the energy landscape. During a memory association process, the energy function value almost monotonically decreases until a convergence. Therefore, it is not possible for a network state to escape from a spurious pattern corresponding to a local minimum.

This property does not hold for the CHN with the CSIGM function. Non-monotonic variations of the energy function value are observed during a memory association process. When the weight matrix is determined by Eq. (2), the energy function satisfies the following inequality [22]:

$$E(\mathbf{s}) \geq -\frac{1}{2}\left(1 - \frac{P}{N}\right) \tag{5}$$

for any pattern $\mathbf{s}$. The equality holds when the pattern is one of the stored patterns.

(d) Network performance The capability of a neural network in an associative memory test can be estimated by the maximum of the load parameter $\alpha = P/N$ for correct recalls. The evaluation of the load parameter reveals that the storage capacity of the CHN with the CSIGN function decreases as K increases [18]. Due to this degradation, it is not appropriate for a multistate associative memory when both the numbers of K and P are large. On the contrary, the CHN with the CSIGM function exhibits less deterioration with an increase of K, although the network capability depends on the degree of its nonlinearity. The properties of the CSIGM function will be discussed in the next section.

3 Complex-sigmoid function

The CSIGM function has been proposed to incorporate continuous nonlinearity into the CSIGM function with discrete nonlinearity [22]. Just by replacing the CSIGN function with the CSIGM one, an improvement of a network performance is expected. Therefore, the CSIGM function can be a standard activation function for complex-valued neural networks, as the sigmoid function is for real-valued neural networks.

The idea of modification to the activation function was motivated by the interpretation of the essence of the CSIGN function. The CSIGN function with K-valued states is defined as follows [11, 18]:

$$\mathrm{CSIGN}(z) = \begin{cases} 1 & 0 \le \arg(z) < \theta_K, \\ \cdots & \cdots \\ e^{ik\theta_K} & k\theta_K \le \arg(z) < (k+1)\theta_K, \\ \cdots & \cdots \\ e^{i(K-1)\theta_K} & (K-1)\theta_K \le \arg(z) < 2\pi, \end{cases} \tag{6}$$

where $z = e^{i\theta}$ is a complex-valued input and $\theta_K = 2\pi/K$ is a scaling factor. This is rewritten as follows:

$$\mathrm{CSIGN}(z) = \exp(ih_K(\arg(z))), \tag{7}$$

where h_K is the multilevel step function:

$$h_K(\theta) = \lfloor \theta/\theta_K \rfloor. \tag{8}$$

The floor function is indicated by $\lfloor\cdot\rfloor$. This is the essence of the CSIGN function except for the scaling between the interval $[0, K)$ and the circle $[0, 2\pi)$. Figure 1(a) shows an example of the multilevel signum function.

Now, we can modify the multilevel step function into the multilevel sigmoid function with smooth continuous nonlinearity leading to a relaxation in the convergence [4, 5, 6, 7, 8, 9]. Unlike the previous studies on multilevel activation functions, we combine the multilevel sigmoid function with circular topology of the complex activation function. The CSIGM function with K-valued states is described as follows:

$$\mathrm{CSIGM}(z) = \exp(if_K(\arg(z))). \tag{9}$$

The multilevel sigmoid function f_K is given by the sum of shifted sigmoid functions as follows:

$$f_K(\theta) = \theta_K \sum_{k=0}^{K-1} g\left(\frac{\theta}{\theta_K} - \left(k + \frac{1}{2}\right)\right), \tag{10}$$

where $g(x) = 1/(1 + \exp(-x/\epsilon))$ with the nonlinearity parameter ϵ controlling the slope of the function. In the limit of $\epsilon \to 0$, $g(x)$ approaches the signum function and thereby $f_K(x) \to h_K(x)$ and $\mathrm{CSIGM}(z) \to \mathrm{CSIGN}(z)$. Figure 1(b) shows an example of the multilevel sigmoid function with $K = 8$. When ϵ is sufficiently small, the function has K stable fixed points near $\theta = k\theta_K$ and K unstable fixed points near $\theta = (k + 1/2)\theta_K$ for $k = 0, \ldots, K-1$.

In the rest of this section, we show some mathematical properties of the CSIGM function.

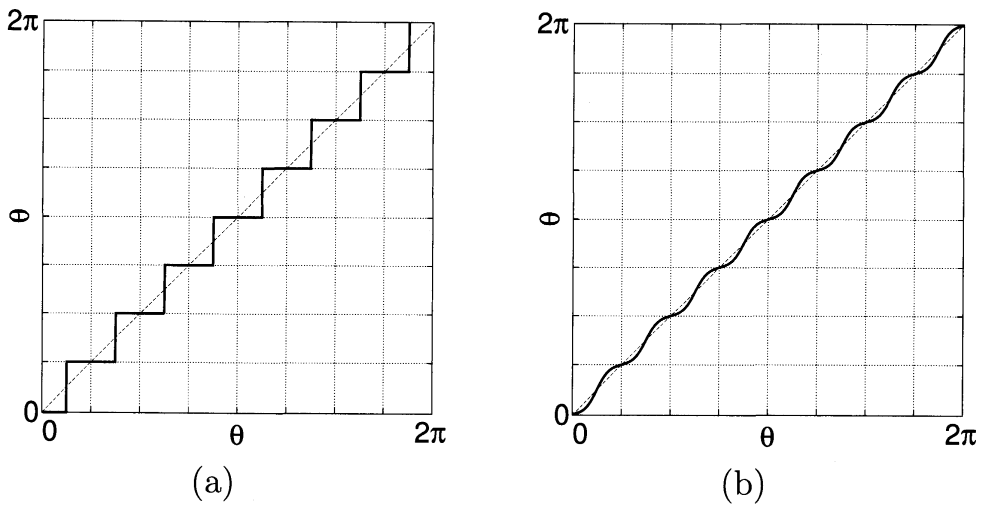


Figure 1: Multilevel functions which are the essences of the complex-valued activation functions: (a) Multilevel signum function with $K = 8$. (b) Multilevel sigmoid function with $K = 8$ and $\epsilon = 0.1$.

(a) Shift transformation The CSIGM function with K levels is nearly invariant under the transformation: $z \longrightarrow z \exp(i\theta_K)$. We consider a shift of the K-level sigmoid function by θ_K as follows:

$$\begin{aligned}
f_K(\theta+\theta_K) &= \theta_K \sum_{k=0}^{K-1} g\left(\frac{\theta+\theta_K}{\theta_K} - \left(k+\frac{1}{2}\right)\right) \\
&= \theta_K \sum_{k=0}^{K-1} g\left(\frac{\theta}{\theta_K} - \left(k-\frac{1}{2}\right)\right) \\
&= \theta_K \left\{\sum_{k=1}^{K} g\left(\frac{\theta}{\theta_K} - \left(k-\frac{1}{2}\right)\right)\right\} \\
&\quad + \theta_K g\left(\frac{\theta}{\theta_K} + \frac{1}{2}\right) - \theta_K g\left(\frac{\theta}{\theta_K} - K + \frac{1}{2}\right) \\
&= f_K(\theta) + \theta_K \left\{ g\left(\frac{\theta}{\theta_K} + \frac{1}{2}\right) - g\left(\frac{\theta}{\theta_K} - K + \frac{1}{2}\right)\right\}. \quad (11)
\end{aligned}$$

Since $0 \leq g(x) \leq e^{-1/\epsilon}$ for $x \leq 0$ and $1 - e^{-1/\epsilon} \leq g(x) \leq 1$ for $x \geq 1$, the two terms in the brackets satisfy the following inequalities, respectively:

$$1 - e^{-1/\epsilon} \leq g\left(\frac{\theta}{\theta_K} + \frac{1}{2}\right) \leq 1, \quad (12)$$

$$0 \leq g\left(\frac{\theta}{\theta_K} - K + \frac{1}{2}\right) \leq e^{-1/\epsilon}, \quad (13)$$

for $0 \le \theta \le 2\pi - \theta_K$. Therefore, we obtain

$$f_K(\theta) + (1 - 2^{-1/\epsilon})\theta_K \le f_K(\theta + \theta_K) \le f_K(\theta) + \theta_K. \tag{14}$$

We have the equality in the multilevel signum function case, i.e. in the limit of $\epsilon \to 0$. When ϵ is sufficiently small, the functional form of the multilevel sigmoid function (and the CSIGM function) for each interval $[k\theta_K, (k+1)\theta_K)$ is almost the same for $k = 0, \ldots, K-1$.

(b) Symmetric property Next, we show a symmetric property of the multilevel sigmoid function and the CSIGM function. Let us consider

$$f_K(2\pi - \theta) = \theta_K \left\{ \sum_{k=0}^{K-1} g\left(\frac{2\pi - \theta}{\theta_K} - \left(k + \frac{1}{2}\right)\right)\right\}. \tag{15}$$

Since $g(-x) = 1 - g(x)$ for any x, it follows

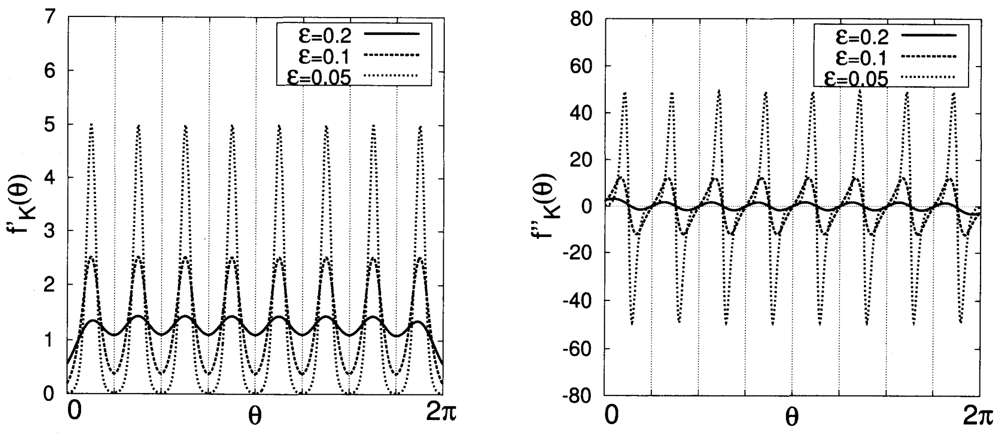
$$\begin{aligned} f_K(2\pi - \theta) &= \theta_K \left\{ \sum_{k=0}^{K-1} 1 - g\left(\frac{\theta}{\theta_K} + \left(k + \frac{1}{2}\right) - K\right)\right\} \\ &= \theta_K \left\{ K - \sum_{k=0}^{K-1} g\left(\frac{\theta}{\theta_K} + \left(k + \frac{1}{2}\right) - K\right)\right\} \\ &= 2\pi - \theta_K \left\{ \sum_{k=0}^{K-1} g\left(\frac{\theta}{\theta_K} - \left(K - k - \frac{1}{2}\right)\right)\right\} \\ &= 2\pi - \theta_K \left\{ \sum_{k=0}^{K-1} g\left(\frac{\theta}{\theta_K} - \left(k + \frac{1}{2}\right)\right)\right\} \\ &= 2\pi - f_K(\theta). \end{aligned} \tag{16}$$

By substituting $\theta = \pi$ into Eq. (16), we get $f_K(\pi) = \pi$. Hence, the multilevel sigmoid function has a point symmetry with respect to the center (π, π). Using Eq. (16), we obtain the following property of the CSIGM function:

$$\mathrm{CSIGM}(\bar{z}) = 1/\mathrm{CSIGM}(z). \tag{17}$$

(c) Monotonicity Using the derivative of the sigmoid function $g'(x) = g(x)(1 - g(x))/\epsilon$, that of the multilevel sigmoid function is described as follows:

$$f'_K(\theta) = \frac{1}{\epsilon} \sum_{k=0}^{K-1} g_k(\theta)(1 - g_k(\theta)), \tag{18}$$



(a) (b)

Figure 2: Multilevel functions which are the essential of complex-valued activation functions: (a) Multilevel step function with $K = 8$. (b) Multilevel sigmoid function with $K = 8$.

where $g_k(\theta) \equiv g(\theta/\theta_K - (k+1/2))$ for $k = 0, \ldots, K-1$. The examples of this function are shown in Fig. 2(a). If $\epsilon \neq 0$, we obtain $f'_K(\theta) > 0$, implying that the multilevel sigmoid function is monotonically increasing. When $\epsilon = 0$, the function is obtained by the sum of the shifted delta functions.

$$\lim_{\epsilon \to 0} f'_K(\theta) = \sum_{k=0}^{K-1} \delta\left(\frac{\theta}{\theta_K} - \left(k + \frac{1}{2}\right)\right). \tag{19}$$

(d) Inflection points The second derivative of the multilevel sigmoid function is written as:

$$f''_K(\theta) = \frac{1}{\epsilon^2 \theta_K} \sum_{k=0}^{K-1} g_k(\theta)(1 - g_k(\theta))(1 - 2g_k(\theta)), \tag{20}$$

which is shown in Fig. 2(b). The function has K inflection points near $\theta = \theta_K(k + 1/2)$ for $k = 0, \ldots, K-1$, where the second derivative changes its sign. The derivative is related to the relaxation time to the stable fixed points.

(e) Bifurcations To investigate the existence and stability of the fixed points, we consider a dynamical system $\theta_{n+1} = f_K(\theta_n)$. The stable (attracting) fixed point points can be found by plotting the converged points after transient periods. Figure 3 shows the bifurcation diagrams for $K = 8$ and

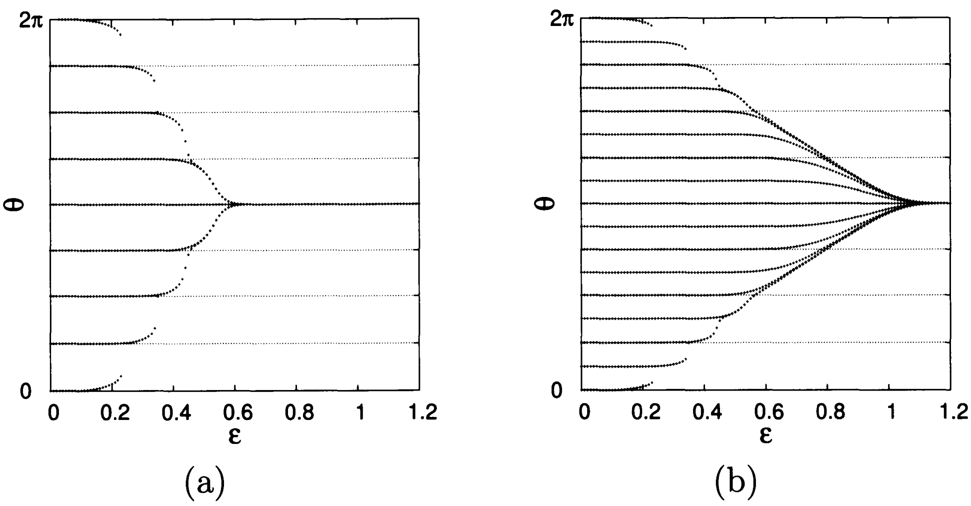


Figure 3: Bifurcation of the CSIGM function for the variation of the nonlinearity parameter ϵ. (a) $K = 8$. (b) $K = 16$.

$K = 16$. We can see that the number of stable fixed points gradually decreases as the nonlinearity parameter ϵ increases. When ϵ is sufficiently large, only the center $\theta = \pi$ remains as the stable fixed point. To guarantee the existence of K stable fixed points, the parameter should be set at a value less than the critical value ϵ^* at which two stable fixed points simultaneously vanish due to the symmetric property. The two stable fixed points are denoted by $\theta = \theta^* \in [0, \theta_K)$ and $\theta = 2\pi - \theta^* \in [2\pi - \theta_K, 2\pi)$. The condition for the critical point can be given as follows:

$$F_1(\theta^*, \epsilon^*) \equiv f_K(\theta^*) - \theta^* = 0, \tag{21}$$

$$F_2(\theta^*, \epsilon^*) \equiv f'_K(\theta^*) - 1 = 0. \tag{22}$$

This is the condition of a tangent bifurcation for the two fixed points. By numerically solving the above two equations with respect to (θ^*, ϵ^*), we get the critical value as $\epsilon^* \sim 0.23093$. This critical value is independent of K. At the bifurcation point, $\theta^*/\theta_K \sim 0.34609$ which is also independent of K.

4 Summary

We have first given an overview of the studies on complex-valued Hopfield network. Several kinds of choices for learning rules and update schemes have been introduced. When the number P of the stored patterns and the number K of states are large, the performance of the network can be improved by replacing the complex-signum function with the complex-sigmoid one including a nonlinearity parameter ϵ [22]. However, the complex-sigmoid function has not yet been fully understood.

In this report, we have investigated some mathematical properties of the complex-sigmoid function and the multilevel sigmoid function. Both functions are nearly invariant under a shift transformation and have symmetric properties. The multilevel sigmoid function is a monotonic function and possess K stable fixed points and K inflection points for a sufficiently small ϵ. As ϵ is increased, the number of stable fixed points gradually decreases. This degeneration occurs due to a sequential disappearance of fixed points via tangent bifurcations. By formulating this condition, we have numerically obtained the critical bifurcation parameter value and the location of the fixed point at the critical point. If the parameter is below the critical value, the existence of K stable fixed points are guaranteed. We have also shown that this critical value is independent of K.

The effect of the nonlinearity parameter on the network capability is still an issue to be explored. The mathematical properties revealed in this report can be useful in analyzing the correlation between the nonlinearity parameter and the network performance.

References

[1] C. M. Bishop, editor. *Pattern recognition and machine learning.* Springer, New York, 2006.

[2] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci USA*, 79:2554–2558, 1982.

[3] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. *Parallel Distributed Processing*, 1:318–362, 1986.

[4] M. Fleisher. The hopfield model with multi-level neurons. In *Proc. AIP Conf.: Neural Inform. Processing Syst.*, pages 278–289, 1987.

[5] H. Rieger. Storing an extensive number of grey-toned patterns in a neural network using multistate neurons. *J. Phys. A: Math. Gen.*, 23:L1273–L1279, 1990.

[6] G. A. Kohring. On the Q-State Neuron Problem in Attractor Neural Networks. *Neural Networks*, 6:573–581, 1993.

[7] J. Si and A. N. Michel. Analysis and synthesis of discrete-time neural networks with multilevel threshold functions. *IEEE Trans. Neural Netw.*, 6(1):105–116, 1995.

[8] J. M. Zurada, I. Cloete, and E. van der Poel. Generalized hopfield networks with multiple stable states. *Neurocomput.*, 13:135–149, 1996.

[9] V. K. Asari. Training of a feedforward multiple-valued neural network by error backpropagation with a multilevel threshold function. *IEEE Trans. Neural. Netw.*, 12(6):1519–1521, 2001.

[10] N. N. Aizenberg, Yu. L. Ivaskiv, and D. A. Pospelov. About one Generalization of the Threshold Function (in russian). *Doklady Akademii Nauk SSSR (the Reports of the Academy of Sciences of USSR)*, 196(6):1287–1290, 1971.

[11] N. N. Aizenberg and I. N. Aizenberg. CNN based on multivalued neuron as a model of associative memory for gray-scale images. In *Proc. 2nd Int. Workshop Cellular Neural Networks Applications*, page 36, 1992.

[12] I. Aizenberg, N. Aizenberg, and J. Vandewalle, editors. *Multi-valued and universal binary neurons: Theory, learning, and applications.* Kluwer Academic Publishers, Boston-Dordrecht-London, 2000.

[13] H. Aoki, R. A-Sadjadi, and Y. Kosugi. Image association using a complex-valued associative memory model. *IEICE Trans. Fundamentals*, 83-A:1824–1832, 2000.

[14] H. Aoki and Y. Kosugi. Characteristics of the complex-valued associative memory model having penalty term. *Electronics and Communications in Japan Part 3*, 83(7):62–70, 2000.

[15] A. Hirose, editor. *Complex-Valued Neural Networks: Theories and Applications*, volume 5 of *Innovative Intelligence.* World Scientific, 2004.

[16] A. Hirose. *Complex-Valued Neural Networks*, volume 32 of *Studies in Computational Intelligence.* Springer-Verlag, 2009.

[17] T. Nitta, editor. *Complex-Valued Neural Networks: Utilizing High-Dimensional Parameters.* Information Science Reference, Pennsylvania, USA, 2009.

[18] S. Jankowski, A. Lozowski, and J. M. Zurada. Complex-valued multistate neural associative memory. *IEEE Trans. Neural Netw.*, 7:1491–1496, 1996.

[19] M. K. Müezzinoğlu, C. Güzeliş, and J. M. Zurada. A new design method for the complex-valued multistate hopfield associative memory. *IEEE Trans. Neural Netw.*, 14(4):891–899, 2003.

[20] G. Tanaka and K. Aihara. Multistate associative memory with parametrically coupled map networks. *Int. J. Bifurcation Chaos*, 15(4):1395–1410, 2005.

[21] D.-L. Lee. Improvements of complex-valued Hopfield associative memory by using generalized projection rules. *IEEE Trans. Neural Netw.*, 17(5):1341–1347, 2006.

[22] G. Tanaka and K. Aihara. Complex-valued multistate associative memory with nonlinear multilevel functions for gray-level image reconstruction. *IEEE Trans. Neural Netw.*, 20(9):1463–1473, 2009.