The Process of Learning Dynamical Systems in Neural Networks and Riddled Basins

Hiroyuki Nakajima

Department of Electronic Engineering and Computer Science,
Faculty of Engineering, Kinki University
1 Takaya-umenobe, Higashi-Hiroshima, 739-21 JAPAN

Abstract

Learning chaotic dynamical systems in neural networks is analyzed as a temporal evolution process in a non-autonomous discrete dynamical system driven by a chaotic teacher signal, defined on the space of connection weights. It is proved to be impossible to derive the condition for the learning converging to the optimal state in general. The proof is based on the stability analysis of the linearized equation about a fixed point corresponding to the optimal state. This impossibility is due to the non-uniformity of chaotic time series and to the fact that a chaotic invariant set is not minimal. It is also pointed out that this non-uniformity can make the basin of the optimal state a very complicated one called a "riddled basin".

1. Introduction

The Learning of a dynamical system in neural networks is the identification of the system referring a time series generated from it as a teacher signal. [1]. Since the learning is successive adjustments of connection weights (i.e., system parameters of neural networks) using a time series presented as a teacher signal, it can be regarded as a non-autonomous dynamical system on the space of connection weights, driven by the teacher signal. The optimal point of the learning is a fixed point of this dynamical system. Hence the convergence condition of the learning can be analyzed as the stability of the fixed point.

From such a point of view, we have been studying analyses of the convergence condition of the learning of dynamical systems in neural networks regarding the learning as a time evolution process in a non-autonomous dynamical system defined on the space of connection weights [1]-[4]. In this article, we discuss the following two points concerning the learning of chaotic time series;

(1) Difficulty of the linear-approximation-based analysis of the convergence of learning chaotic times series

(2) A riddled basin of the optimal state of the learning of chaos.

2. Mathematical Formulation of the Learning of Dynamical Systems

We deal with the successive-type back-propagation (BP) learning of an m-dimensional discrete dynamical system,

$$\boldsymbol{q}(k+1) = f(\boldsymbol{q}(k)) \qquad (\boldsymbol{q} \in \boldsymbol{R}^m). \tag{1}$$

Where f is a continuous map on $\boldsymbol{R}^n$ and k denotes continuous time. Referring input-output patterns, $(\boldsymbol{q}(k), f(\boldsymbol{q}(k)))$, as a teacher signal, a neural network, described by the following equation, learns the above system.

$$\boldsymbol{x}(k+1) = \tilde{f}(\boldsymbol{x}(k), \boldsymbol{w}) \qquad (\boldsymbol{x} \in \boldsymbol{R}^m). \tag{2}$$

[1]Strictly speaking, we should say "to identify a dynamical system by learning the time series generated from it". In this article, however, we will say "to learn a dynamical system" for the sake of simplicity.

Where $\boldsymbol{w}$ is an n-dimensional vector representing connection weights of the neural network.

The BP learning of successive type is described by a kind of gradient-descent method,

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \eta \nabla_{\boldsymbol{w}} E(\boldsymbol{w}(k), k). \tag{3}$$

Where η is a positive number called a learning constant, and E is a function evaluating the error between the output of the neural network and the teacher signal, which function is defined by,

$$E(\boldsymbol{w}, k) = \frac{1}{2} \| f(\boldsymbol{q}(k)) - \tilde{f}(\boldsymbol{q}(k), \boldsymbol{w}) \|^2. \tag{4}$$

Since the teacher signal $\boldsymbol{q}(k)$ varies as the time k, the error function E depends on k explicitly. Thus the learning equation (3) is a non-autonomous dynamical system of the connection weights $\boldsymbol{w}$, and the optimal point $\boldsymbol{w}^*$, for which $E(\boldsymbol{w}^*, k) = 0$ for every time k, is a fixed point of this dynamical system. If $\boldsymbol{w}^*$ is asymptotically stable as a fixed point of (3), local convergence of the learning to the optimal point is guaranteed. This formulation can be applied to more general learning by a successive gradient-descent method (3) in a system (2) with adjustable parameters as well as the BP-learning in neural networks.

3. Analysis of the Convergence Condition of the Learning of Chaos

In this section, we analyze the learning of chaotic time series based on the mathematical formulation introduced in the previous section.

We denote the non-autonomous discrete dynamical systems defined by Eqs. (3) and (4) by,

$$\boldsymbol{w}(k+1) = g(\boldsymbol{w}(k), k). \tag{5}$$

Where the explicit dependence of the function g on time k is due to the fact that the teacher signal $\boldsymbol{q}(k)$ is time varying governed by Eq. (1). To indicate this time dependence more explicitly, we also describe the system (5) by,

$$\boldsymbol{w}(k+1) = \bar{g}(\boldsymbol{w}(k), \boldsymbol{q}(k)). \tag{6}$$

Now we consider the stability of the fixed point $\boldsymbol{w}^*$ of this dynamical system based on the linear-approximation analysis. Linear approximation is widely used for stability analysis of a fixed point in dynamical systems. An important theoretical back-ground of this analysis is the following lemma [5]:

Lemma Let $\boldsymbol{w}^*$ be a fixed point of a discrete dynamical system,

$$\boldsymbol{w}(k+1) = g(\boldsymbol{w}(k), k) \quad (\boldsymbol{w} \in \boldsymbol{R}^n). \tag{7}$$

If the null solution, $\boldsymbol{w} = \mathbf{o}$, of the linearized equation around the fixed point,

$$\boldsymbol{w}(k+1) = D_{\boldsymbol{w}} g(\boldsymbol{w}^*, k) \boldsymbol{w}(k), \tag{8}$$

is uniformly asymptotically stable, then the fixed point $\boldsymbol{w}^*$ is also uniformly asymptotically stable. □

Here uniformly asymptotic stability of a solution, $\boldsymbol{w}^*$, of (7) is defined as follows:

Definition 1 Let $\boldsymbol{w}(\boldsymbol{w}_0, k_0, k)$ be a solution of (7) which starts $\boldsymbol{w}_0$ at an initial time $k = k_0$. We say a solution $\boldsymbol{w}^*$ is uniformly asymptotically stable if it satisfies the following two conditions [6];

(i) For every positive number ϵ, there exists a positive number δ such that if $\|\boldsymbol{w}_0 - \boldsymbol{w}^*\| < \delta$, then for every integer k_0 and every $k \geq k_0$, $\|\boldsymbol{w}(\boldsymbol{w}_0, k_0, k) - \boldsymbol{w}^*\| < \epsilon$ is satisfied (uniform stability).

(ii) There exists a positive number δ such that for any positive number ϵ there exists an integer $K(\epsilon)$, if $\|\boldsymbol{w}_0 - \boldsymbol{w}^*\| < \delta$ then for every integer k_0 and for any $k \geq k_0 + K(\epsilon)$, $\|\boldsymbol{w}(\boldsymbol{w}_0, k_0, k) - \boldsymbol{w}^*\| < \epsilon$ is satisfied. □

An important point of this definition is that the time $K(\epsilon)$ in the condition (ii), which the orbits starting from the δ-neighborhood of the fixed point $\boldsymbol{w}^*$ take before they converge into the ϵ-neighborhood of $\boldsymbol{w}^*$, does not depend on the initial time k_0. Thus the term "uniform" means the uniformity with respect to time k. When the function $\bar{g}(\boldsymbol{w}, \boldsymbol{q}(k))$ is periodic with respect to time k, it is easy to see that this uniformity is satisfied. In such a case, the lemma is valid even if the "uniformly asymptotically stable" in the above definition is replaced with "asymptotically stable". Therefore if $\boldsymbol{q}(k)$ in Eq. (6) is periodic (so, if it is a fixed point, of course), the asymptotic stability of the fixed point $\boldsymbol{w}^*$ is guaranteed by the asymptotic stability of the null solution, $\boldsymbol{w} = \mathbf{o}$, of the variational equation linearized around $\boldsymbol{w}^*$. In such a case, the stability analysis is not so difficult and the asymptotic stability of the null solution is proved for many cases [1, 3]. In other words, local convergence of the learning of periodic orbits is guaranteed in general.

However, when $\boldsymbol{q}(k)$ is chaotic as is discussed in this article, the situation is completely different. That is, the null solution of the linear approximated system,

$$\boldsymbol{w}(k+1) = D_{\boldsymbol{w}}\bar{g}(\boldsymbol{w}^*, \boldsymbol{q}(k))\boldsymbol{w}(k), \tag{9}$$

is not uniformly asymptotically stable in general. This is expressed by the following theorem:

Theorem 1 Let $\boldsymbol{q}(k)$ be a dense orbit contained in a chaotic attractor A_T that contains at least one periodic orbit with a period less than n/m. Then the null solution of the linearized system (9) is not uniformly asymptotically stable. □

Thus the uniformly asymptotic stability of the linear approximated system is not guaranteed in general. However, non-uniformly asymptotic stability is guaranteed in general as is shown in the following theorem:

Theorem 2 Suppose $\{\boldsymbol{q}(k)\}$ is an orbit contained in a chaotic attractor A_T and there exists a positive integer n' such that $f^{n'}$ is ergodic with respect to an invariant probabilistic measure P of A_T. And suppose

$$P(\ \{\ \boldsymbol{q} \mid r(\boldsymbol{q}) < 1\ \}\ ) > 0 \tag{10}$$

is satisfied. Where $r(\boldsymbol{q})$ is defined by

$$r(\boldsymbol{q}) = \|D_{\boldsymbol{w}}\bar{g}(\boldsymbol{w}^*, f^{n'-1}(\boldsymbol{q}))\cdots D_{\boldsymbol{w}}\bar{g}(\boldsymbol{w}^*, \boldsymbol{q})\|.$$

Besides we assume the learning constant η is sufficiently small so that $|\lambda^{(l)}(k)| < 1$ is satisfied for all $k \in \boldsymbol{Z}$ and $l = 1, \ldots, m$. Where

$$\lambda^{(l)}(k) = 1 - m\eta\|\boldsymbol{a}^{(l)}(k)\|^2$$

and $\boldsymbol{a}^{(l)}(k)$ is an n-dimensional vector defined by,

$$\boldsymbol{a}^{(l)}(k) = (a_{k1}^{(l)}, \ldots, a_{kn}^{(l)})^T,\ a_{ki}^{(l)} = \frac{\partial[\tilde{f}_l(\boldsymbol{q}(k), \boldsymbol{w}) - f_l(\boldsymbol{q}(k))]}{\partial w_i}\ (l = 1, \ldots, m, i = 1, \ldots, n, k \in \boldsymbol{Z}).$$

Then the null solution of (9) is asymptotically stable with probability one with respect to the initial value of $\boldsymbol{q}(k)$. □

For the proofs of these two theorems, see Refs. [1, 3]. Here we will explain the meaning of them intuitively.

A chaotic orbit is dense in a chaotic invariant set A_T, and the set contains unstable periodic orbits (UPO) innumerably [7]. So the conditions of the above theorems are satisfied in general.

A dense chaotic orbit in A_T has chances to approach the UPOs arbitrarily closely although it can never converge to such an unstable orbit. And the closer it approaches to a UPO, the longer it stays around there. In terms of learning, therefore, the closer the initial value $\boldsymbol{q}(k_0)$ is to the UPO at an initial time k_0, the longer the time is, during which the presented teacher signal patterns are restricted to those around the UPO. Therefore even if the learning with a certain chaotic teacher signal converges to the optimal point, the time it takes is not uniform with respect to the initial time k_0. We will explain this situation using a simple example of a learning problem as follows:

Here we consider the learning of the Logistic map,

$$q(k+1) = w^* q(k)(1-q(k)) \quad (q \in [0,1]), \tag{11}$$

in a learning system with a one-dimensional adjustable parameter w,

$$x(k+1) = wx(k)(1-x(k)) \quad (x \in \boldsymbol{R}). \tag{12}$$

Where we set $w^* = 4$. Since an error evaluating function is defined by,

$$E(\boldsymbol{w}, k) = \frac{1}{2}(w(k) - w^*)^2 [q(k)(1-q(k))]^2, \tag{13}$$

the learning based on gradient-descent method minimizing this function is described by,

$$w(k+1) = w(k) - \eta(w(k) - w^*)[q(k)(1-q(k))]^2. \tag{14}$$

This equation itself is affine with respect to w, so the linearized system about w^* of it is a system of a variable v changed from w by $v \equiv w - w^*$ without approximation,

$$v(k+1) = (1 - \eta[q(k)(1-q(k))]^2)v(k). \tag{15}$$

Since a chaotic orbit $q(k)$ of the Logistic map is restricted to an open interval $(0,1)$, the coefficient $1 - \eta[q(k)(1-q(k))]^2$ in the right hand side of Eq. (15) is less than unity in its magnitude if the learning constant η is sufficiently small. Thus we can expect the null solution, $v = 0$, of the linear equation (15) is asymptotically stable. It is Theorem 2 that guarantees this in general.

However, since $q(k)$ is dense in a closed interval $[0,1]$, it can approach a fixed point $q = 0$ arbitrarily closely. The closer $q(k_0)$ is to 0 at an initial time k_0, the longer the time is, during which the coefficient in the right hand side of Eq. (15) remains around unity, and thus the longer the convergence time of $v(k)$ to 0 becomes. Hence there does not exist a convergence time $K(\epsilon)$ that does not depend on the initial time k_0 as is required in the condition (ii) of Definition 1. Therefore the null solution cannot be uniformly asymptotically stable.

Hondou et al. [8]–[10] have pointed out that the short-term correlation caused by the stay of chaotic orbits around a UPO accelerates the learning of chaos. Furthermore, they also have showed that due to such a correlation, when a particle in a periodic potential is excited by a chaotic signal, the transition probability of the particle from a potential valley to another becomes asymmetric even if the signal has a uniform invariant measure and the power spectral density of it is white [8]–[10]. These results imply the possibility of utilizing the temporal non-uniformity of chaos in learning, while in this article we have discussed a negative result, that is, this non-uniformity of chaos makes it hard to analyze the condition for convergence of the learning.

In relation to the non-uniformity of chaos, we should emphasize the difference between chaotic orbits and almost periodic orbits. To explain this difference, first we introduce the definition of minimal sets.

Definition 2 If a closed invariant set S of a discrete dynamical system contains no closed invariant proper subset, then S is called a minimal set [11, 12]. □

For example, clearly a periodic attractor is minimal. A chaotic attractor generally contains UPOs, which are closed invariant proper subsets, hence it is not a minimal set. An almost periodic orbit is also aperiodic. However its closure, namely a torus, contains no UPOs. Hence a torus is a minimal set and an almost periodic orbit is temporally uniform unlike a chaotic orbit. Therefore uniformly asymptotic stability of the null solution of the linearized system (9) of the learning of an almost periodic orbit is guaranteed in general [3]. Minimality is an important property that characterizes the difference between chaos and torus as well as Lyapunov exponents or power spectrum.

4. A Riddled Basin of the Optimal State of Learning

In the previous section, we considered the learning of chaos in neural networks to be a non-autonomous dynamical system on the space of connection weights driven by a chaotic teacher signal. On the other hand, we can regard the learning as a time evolution process in an extended system of the system of connection weights combined with the system (1), which generates the chaotic teacher signal. In this section, we show that the optimal state of the learning is a chaotic invariant set restricted to a low-dimensional subspace in the extended system, and that due to the non-minimality of the invariant set caused by the UPOs contained in it, the basin of the set can be a complicated one called a riddled basin [13].

We deal with the following extended system, which is a combination of the learning system (6) with a chaotic system (1), which generates the teacher signal,

$$\begin{aligned} \boldsymbol{w}(k+1) &= \bar{g}(w(k), \boldsymbol{q}(k)), \\ \boldsymbol{q}(k+1) &= f(\boldsymbol{q}(k)). \end{aligned} \tag{16}$$

We denotes the chaotic attractor that contains the teacher signal $\boldsymbol{q}$ by $A_q \subset \boldsymbol{R}^m$. The extended system (16) is an $m \times n$-dimensional autonomous system defined on $A_q \times \boldsymbol{R}^n$. The optimal state of the learning is an m-dimensional subspace, $A = A_q \times \{\boldsymbol{w}^*\}$, which is a chaotic invariant set. Hence the convergence of the leaning can be analyzed as the stability of this chaotic invariant set A.

A typical example of such a chaotic invariant set restricted to a subspace is a state of chaotic synchronization. For example, chaotic synchronization of master-slave type proposed by Pecora & Carroll [14] is described by a continuous dynamical system,

$$\dot{\boldsymbol{x}}(t) = \phi(\boldsymbol{x}(t), \boldsymbol{y}(t)), \quad \dot{\boldsymbol{y}}(t) = \psi(\boldsymbol{x}(t), \boldsymbol{y}(t)), \quad \dot{\boldsymbol{z}}(t) = \psi(\boldsymbol{x}(t), \boldsymbol{z}(t)). \tag{17}$$

Where $\boldsymbol{x}$ is an n-dimensional vector and $\boldsymbol{y}$ and $\boldsymbol{z}$ are m-dimensional vectors. We suppose $\boldsymbol{x}$ and $\boldsymbol{y}$ are oscillating chaotically with interaction described by the first and the second equations of (17). Since the forms of the second and the third equations of (17) are identical, there exists a solution satisfying $\boldsymbol{y}(t) \equiv \boldsymbol{z}(t)$. That is, it is possible that $\boldsymbol{y}$ and $\boldsymbol{z}$ synchronize chaotically. Furthermore $\boldsymbol{y}$ affects $\boldsymbol{z}$ via the interaction with $\boldsymbol{x}$, while $\boldsymbol{z}$ does not affect other variables but only receives external forcing. Therefore this is called the master ($\boldsymbol{y}$) and slave ($\boldsymbol{z}$) type chaotic synchronization.

With a change of variable from $\boldsymbol{z}$ to $\boldsymbol{w} = \boldsymbol{z} - \boldsymbol{y}$, the third equation of (17) is transformed into,

$$\dot{\boldsymbol{w}}(t) = \psi(\boldsymbol{x}(t), \boldsymbol{w}(t) + \boldsymbol{y}(t)) - \psi(\boldsymbol{x}(t), \boldsymbol{y}(t)). \tag{18}$$

The state of chaotic synchronization is defined by $\boldsymbol{w} = \mathbf{o}$, and this is a chaotic invariant set restricted to an $(n+m)$-dimensional subspace, $\{\ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) \mid \boldsymbol{w} = 0\ \}$, of the $(n+2m)$-dimensional system composed of the first and the second equations of (17) and Eq. (18).

Consequently, the learning of chaos and chaotic synchronization can be formulated in the basically same form.

Alexander et al. [13] discovered that the basin of such a chaotic invariant set restricted to a subspace can be a very complicated one that is eroded everywhere by another attractor's basin. Such a basin is called a riddled basin, which is defined as follows:

Definition 3 If an invariant set A of a dynamical system is an attractor in the sense of Milnor, and if for every point x in its basin $B(A)$ and for every open neighborhood $U(x)$ of x, $U(x) \cap B(A)^C$ has positive measure, then $B(A)$ is called a riddled basin. □

Where the basin $B(A)$ is defined as the set of points whose ω-limit set is contained in A. An attractor in the sense of Milnor is defined as follows [15]:

Definition 4 A compact invariant set A of a dynamical system is called an attractor in the sense of Milnor if the following two conditions are satisfied:

(1) The basin $B(A)$ has positive Lebesgue measure.

(2) There exists no proper subset A' of A, whose basin $B(A')$ is equal to $B(A)$ up to a set of zero measure. □

An attractor is ordinarily defined as an invariant set whose basin is a neighborhood of it. Milnor's definition is an extended version of conventional definitions, which regards an invariant set whose basin has positive measure as an attractor even if it is not a neighborhood. The basin $B(A)$ of an invariant set A is a riddled basin implies it is everywhere eroded by its compliment set to arbitrarily near the points of A and is not a neighborhood of A, although it has positive measure. An attractor A with a riddled basin attracts many orbits starting from its neighborhood, but the set of initial points the orbits starting from where do not converge to A is also dense and has positive measure.

Now we introduce an example of a system for learning chaos in which a riddled basin is observed [3, 4]. Here we consider the learning of the tent map,

$$q(t+1) = T(q(t)) = \begin{cases} 2q(t) & (0 \le q \le \frac{1}{2}), \\ 2 - 2q(t) & (\frac{1}{2} < q \le 1), \end{cases} \tag{19}$$

in a learning system with one-dimensional adjustable parameter w,

$$x(t+1) = \tilde{f}(x(t), w) \quad = \quad (g(w)+2) \times \begin{cases} x(t) & (0 \le x \le \frac{1}{2}), \\ 1 - x(t) & (\frac{1}{2} < x \le 1), \end{cases} \tag{20}$$

$$g(w) \quad = \quad \begin{cases} -\frac{1}{3}(w+4) & (w < -1), \\ w & (|w| \le 1), \\ \frac{1}{3}(-w+4) & (w > 1). \end{cases} \tag{21}$$

The learning is described by the following gradient descent procedure,

$$w(t+1) = w(t) - \frac{1}{4}\eta g(w(t)) g'(w(t)) T(q(t))^2. \tag{22}$$

It is proved that a chaotic invariant set, $A = \{\ (q, w) \mid q \in [1, 0], w = 0\ \}$, in this system is an attractor in the sense of Milnor and its basin is a riddled basin [3, 4]. Figure 1 is a numerical example of the riddled basin. The basin $B(A)$ (the white region) is riddled by its compliment set.

It is nothing but the non-minimality of chaos caused by the UPOs contained in it, which makes the analysis of the convergence of learning chaos based on linear approximation difficult, that generates such a strange structure of the basin.

In general, the stability of the chaotic invariant set A in the system (16) can be evaluated by an exponent called a normal Lyapunov exponent, which is an expansion rate of orbits along the $\boldsymbol{w}$-direction,

$$\lambda = \lim_{k\to\infty} \frac{1}{k} \log \|\boldsymbol{v}(k)\|. \tag{23}$$

Where $\boldsymbol{v}(k)$ is a solution of the linearized equation of (16) about $\boldsymbol{w}$-component,

$$\boldsymbol{v}(k+1) = D_{\boldsymbol{w}}\bar{g}(\boldsymbol{w}^*, \boldsymbol{q}(k))\boldsymbol{v}(k), \tag{24}$$

and $\boldsymbol{q}(k)$ is a chaotic orbit dense in A_q.

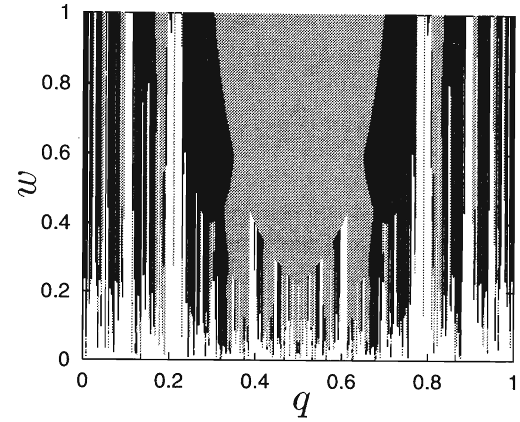

Fig. 1 A riddled basin of A in (22) ($\eta = 22.0$. The white region is the basin of A.)

It is known that if the exponent λ is negative, the null solution of the linearized system (24) is asymptotically stable and the measure of the basin of A is positive in general. However, λ is an expansion rate averaged along the chaotic orbit, $\boldsymbol{q}(k)$, which conserves the natural invariant measure on A. Hence even if it is negative, the expansion rate calculated in the neighborhood of the UPOs contained in A, which UPOs do not conserve the natural invariant measure, may be positive. In such a case, there may exist a route from the neighborhood of the UPOs to another attractor and the basin $B(A)$ may be eroded. Moreover, since the stable manifolds of the UPOs are dense in A in general, the erosion occurs everywhere arbitrarily near A. This is the mechanism of the generation of a riddled basin.

The riddled basin in learning is a manifestation of the spatial nonuniformity (non-minimality) of the learning of chaos described by an autonomous system (16), which appears in Theorem 1 as the temporal nonuniformity that makes it hard to analyze the convergence condition of the learning described by a non-autonomous system (6).

In the analysis of chaotic synchronization, the normal Lyapunov exponent (23) is called a conditional Lyapunov exponent (or a transversal Lyapunov exponent). In general, the synchronization state is asymptotically stable if this exponent is negative [14]. Does this fact contradict the existence of a riddled basin? The answer is "No". The stability analysis using Lyapunov exponents is based on linear approximation. Therefore, in the system (18) for example, if the initial value $\boldsymbol{w}(0)$ is sufficiently near the synchronization state, $\boldsymbol{w} = \mathbf{o}$, then the synchronization is guaranteed by the negative Lyapunov exponents. However, the sufficiency depends on the initial conditions of signals, $\boldsymbol{x}$ and $\boldsymbol{y}$. And there cannot exist such a value, the orbits with a initial value $\boldsymbol{w}(0)$ less than which in its magnitude are guaranteed to converge the synchronization state for all initial conditions of $\boldsymbol{x}$ and $\boldsymbol{y}$ if they are chaotic. That is, the fact that $\boldsymbol{w}^*$ cannot be uniformly asymptotically stable, as is shown in Theorem 1, generates the riddled basin.

5. Conclusion

The results are summarized as follows: When we analyze the convergence condition to the optimal state of the learning of chaotic dynamical systems in neural networks as the stability of a fixed point of a dynamical system defined on the space of connection weights driven by chaotic forcing, the linear approximation-based analysis is difficult, and the difficulty is caused by the non-minimality of chaotic invariant set containing UPOs and is related to riddled basins. We intuitively explained these results including the relation to chaotic synchronization, which we could not contained in Refs. [1, 2, 3], while one can find the mathematical discussion about it in these references.

This paper was presented in part at the 9th SICE Symposium on Decentralized Autonomous Systems.

References

[1] H. Nakajima and Y. Ueda, "An Analysis of the Process of Learning Discrete Dynamical Systems by Back Propagation Method Using the Theory of Non-autonomous Difference Equations", *Trans. IEICE*, **J77-D-II** No. 11, pp. 2279-2286 (1994)

[2] H. Nakajima and Y. Ueda, "An Analysis of the Process of Learning Continuous Dynamical Systems in Neural Networks Using the Theory of Non-autonomous Difference Equation", *Trans. IEICE*, **J78-D-II** No. 12, pp. 1890-1898 (1995)

[3] H. Nakajima, "Nonlinear Analyses of the Processes of Learning Dynamical Systems in Neural Networks", Doctoral Thesis, Kyoto University (1995)

[4] H. Nakajima and Y. Ueda, "Riddled Basins of the Optimal States in Learning Dynamical Systems", *Physica* **D**, **99**, pp. 35-44 (1996)

[5] J. L. Willems: "Stability Theory of Dynamical Systems", Thomas Nelson & Sons (1970)

[6] S. Sugiyama: "Sabun-Bibun Houteishiki", Kyoritsu Shuppan (1971)

[7] R. L. Devaney: "An Introduction to Chaotic Dynamical Systems", Benjamin/Cummings Publishing, Inc. (1986)

[8] T. Hondou and Y. Sawada, "Analysis of Learning Processes of Chaotic Time Series by Neural Networks", *Prog. Theor. Phys.* **91**, pp. 397-402 (1994)

[9] T. Hondou and Y. Sawada, "Dynamical Behavior of a Dissipative Particle in a Periodic Potential Subject to Chaotic Noise: Retrieval of Chaotic Determinism with Broken Parity", *Phys. Rev. Lett.*, **75**, pp. 3269-3272 (1995)

[10] T. Hondou, M. Yamamoto, Y. Sawada, and Y. Hayakawa, "Effect of Time-correlation of Input Patterns on the Convergence of On-line Learning", *Phys. Rev.* **E 53**, pp. 4217-4220 (1996)

[11] T. Saito: "Isou Rikigaku", Kyoritsu Shuppan (1971)

[12] N. Aoki and K. Shiraiwa: "Rikigakukei To Entropy", Kyoritsu Shuppan (1985)

[13] J. C. Alexander, J. A. Yorke, Z. You and I. Kan, "Riddled Basins", *International Journal of Bifurcation and Chaos*, **2**, No. 4, pp. 795-813 (1992)

[14] T. L. Carroll and L. M. Pecora, "Synchronizing Chaotic Circuits", *IEEE Trans. Circuits Syst.*, **CAS-38**, No. 4, pp. 453-456 (1991)

[15] J. Milnor, "On the Concept of Attractor", *Commun. Math. Phys.*, **99**, pp. 177-195 (1985)