PAPER Special Section of Selected Papers from the 8th Karuizawa Workshop on Circuits and Systems

An Analysis on Minimum Searching Principle of Chaotic Neural Network

Masaya OHTA†, Kazumichi MATSUMIYA†*, *Student Members*, Akio OGIHARA†, Shinobu TAKAMATSU†, *and* Kunio FUKUNAGA†, *Members*

SUMMARY This article analyzes dynamics of the chaotic neural network and minimum searching principle of this network. First it is indicated that the dynamics of the chaotic neural network is described like a gradient descent, and the chaotic neural network can roughly find out a local minimum point of a quadratic function using its attractor. Secondly It is guaranteed that the vertex corresponding a local minimum point derived from the chaotic neural network has a lower value of the objective function. Then it is confirmed that the chaotic neural network can escape an invalid local minimum and find out a reasonable one.

key words: *chaos, neural network, minimum searching problem, attractor*

1. Introduction

Recently the chaotic neural network is studied from the viewpoint of a minimum searching machine. Nozawa [1] has proposed a new method of solving the traveling salesman problem (TSP) using the chaotic neural network and has shown solving ability experimentally. Chen and Aihara [2] have proposed a chaotic simulated annealing and confirmed the ability of the chaotic neural network. However the mechanism of the minimum searching by the chaotic behavior is not clear.

We analyze dynamics of the chaotic neural network and minimum searching principle of this network. First the chaotic neural network is defined and its behavior is considered theoretically and experimentally. As a result we prove that the dynamics of the chaotic neural network is described like a gradient descent. Then we confirm that the chaotic neural network can roughly find out a local minimum point of a quadratic function using its attractor. Secondly we guarantee that the vertex corresponding a local minimum point derived from the chaotic neural network has a lower value of the objective function. Then we indicate that the chaotic neural network can escape an invalid local minimum and find out a reasonable one.

2. Chaotic Neural Network

In this section a chaotic neural network is defined. Its behavior is considered theoretically and experimentally. As a result we will prove that the dynamics of the chaotic neural network is described like a gradient descent. Furthermore we will say that the chaotic neural network can roughly find out a local minimum point of a quadratic function using its attractor.

2.1 Definition of Chaotic Neural Network

The chaotic neural network in this article is derived from the differential equations of the Hopfield's model [3]. It has proposed by Nozawa [1]. In the Hopfield's model, the behavior of neuron i ($i = 1, 2, \ldots, M$) is defined as follows [3] :

$$\frac{du_i(t)}{dt} = -\frac{u_i(t)}{R} + \sum_{j=1}^{M} T_{ij} v_j(t) + I_i \tag{1}$$

and

$$v_i(t) = S(u_i(t)) = \frac{1}{1+\exp(-u_i(t)/\alpha)}, \tag{2}$$

where $u_i(t)$ is the input of neuron i at continuous time t, $v_i(t)$ is the output of neuron i, T_{ij} is the synaptic connection of neuron j to neuron i, I_i is the threshold value of neuron i, R (> 0) is the damping constant of the input, $S(\cdot)$ is the sigmoidal function and α (> 0) is the gain constant of the function S.

The chaotic neural network is defined as the difference equation version of Eqs. (1) and (2) by Euler's method with the difference step Δt.

$$u_i(n) = \Delta t \sum_{j=1}^{M} T_{ij} \sum_{k=0}^{n} \left(1 - \frac{\Delta t}{R}\right)^k v_j(n-1-k) + RI_i$$

and (3)

$$v_i(n) = S(u_i(n)) = \frac{1}{1+\exp(-u_i(n)/\alpha)}, \tag{4}$$

where n is the discrete time. It is assumed that $u_i(0) = 0$ and n is an enough large number at the change from Eq. (1) to Eq. (3).

An internal buffer of neuron i at the discrete time n is defined as

$$p_i(n) = \frac{\Delta t}{R} \left\{ \sum_{k=0}^{n-1} \left(1 - \frac{\Delta t}{R}\right)^k v_i(n-1-k) \right\}. \tag{5}$$

Manuscript received July 3, 1995.
Manuscript revised October 5, 1995.

†The authors are with the College of Engineering, University of Osaka Prefecture, Sakai, 593 Japan.
*Presently, with Tokyo Institute of Technology.

From Eq. (5), Eq. (3) is rewritten as

$$u_i(n) = R\left(\sum_{j=1}^{M} T_{ij}p_j(n) + I_i\right). \tag{6}$$

Then, Eq. (5) is described as follows:

$$\begin{aligned} p_i(n+1) &= \left(1 - \frac{\Delta t}{R}\right)p_i(n) + \frac{\Delta t}{R}v_i(n) \\ &= p_i(n) + \frac{\Delta t}{R}\{v_i(n) - p_i(n)\}. \end{aligned} \tag{7}$$

The behavior of the chaotic neural nètwork, therefore, is described as follows from Eqs. (7), (4) and (6):

$$p_i(n+1) = p_i(n) + \epsilon\Delta_i \quad \text{and} \tag{8}$$

$$\Delta_i = \frac{1}{1+\exp\left\{-\frac{R}{\alpha}\left(\sum_{j=1}^{M} T_{ij}p_j(n) + I_i\right)\right\}} - p_i(n), \tag{9}$$

where $\epsilon = \Delta t/R$. β is defined as $\beta = \frac{\alpha}{RT}$, $T = -T_{ii} > 0$ for the latter.

2.2 Dynamics of Chaotic Neural Network

In this subsection, we investigate a behavior of the chaotic neural network.

Figure 1 (a) shows the conceptual graph of Eq. (9) with $T_{ij} = 0$ $(i \neq j)$ and $0 < -I_i/T_{ii} < 1$. The point 'Z' (illustrated by ∘) corresponds to where the sign of Δ_i changes. At the point 'C' (illustrated by •), the sigmoidal function is 0.5 so that the input for the neuron i is zero; $\sum_j T_{ij}p_j + I_i = 0$. If β is small enough, the point 'C' becomes close to the point 'Z'. Figure 1 (b) shows the integral calculus of $-\Delta_i$. This curve is drawn on the two parabolas corresponding to l_H and l_L respectively. The bottom of this curve corresponds the point 'Z'. From Eq. (8), it is considered that the behavior of $p_i(n)$ is by the gradient descent method, so that the state $p_i(n)$ goes down the point 'Z', approximately the point 'C', on the curve in Fig. 1 (b). An example of dynamics is shown in Fig. 2. This is a return map of $p_i(n)$. The horizontal axis and the vertical axis show $p_i(n)$ and $p_i(n+1)$ respectively, and the solid line shows the trajectory of $p_i(n)$ according to Eqs. (8) and (9), where $\epsilon = 0.26, \alpha/R = 0.006$ and $-I_i/T_{ii} = 0.2$. The point 'C' in Fig. 2 corresponds to 'C' in Fig. 1 (b) and it's close to the bottom 'Z' of the integral calculus. It is obvious from the figure that $p_i(n)$ never stop at the bottom and it wanders around the bottom non-periodically, that is chaotic. We may say that there is an attractor of the chaotic dynamics around the bottom in Fig. 1 (b).

Next we consider a linear dynamical system which is defined by exchanging Δ_i for $\overline{\Delta}_i = \sum_{j=1}^{M} T_{ij}p_j(n) + I_i$ in Eqs. (8) and (9). $\overline{\Delta}_i$ is extracted and defined from the index of the exponential function of Eq. (9). Figure 1(c) shows the graph of $\overline{\Delta}_i$. $\overline{\Delta}_i$ is a linear function of p_i so that it is drawn by a straight line. Then $\overline{\Delta}_i$ intersects to p_i-axis at the point 'C' as same as 'C' of Fig. 1 (a). Because at the point 'C' in Fig. 1 (a) where the sigmoidal function is 0.5, the index of the exponential function of $S(u_i(n))$ is equal to zero, that is $u_i(n) = R\overline{\Delta}_i = 0$, from Eqs. (4) and (6). Thus 'C' of Fig. 1 (c) is equivalent to Fig. 1 (a)'s 'C'. Figure 1 (d) shows the integral calculus of $-\overline{\Delta}_i$. It's an exact quadratic function and it has a local minimum at the point 'C'. The behavior of the linear dynamical system is convergence to this local minimum. In other words, the local minimum of the quadratic function is an asymptotically stable point for the linear system. In Fig. 1 (b), if β is small enough, the sigmoidal function is close to the step function and

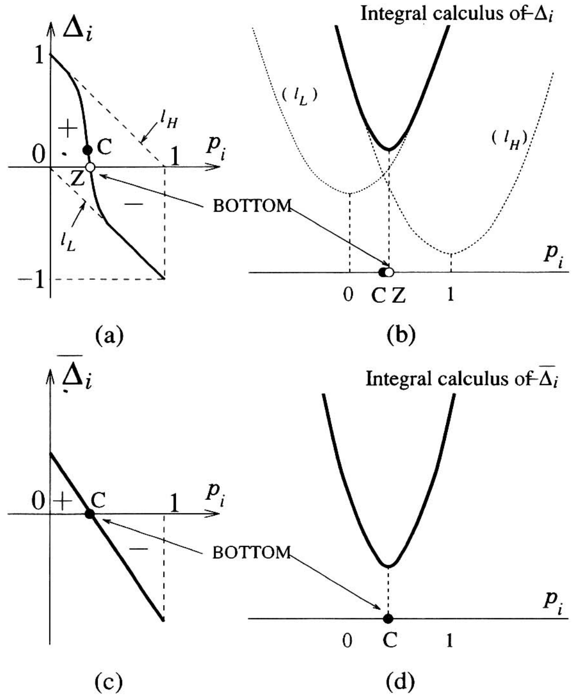

Fig. 1 Graphs of Δ_i and $\overline{\Delta}_i$, and their integral calculuses.

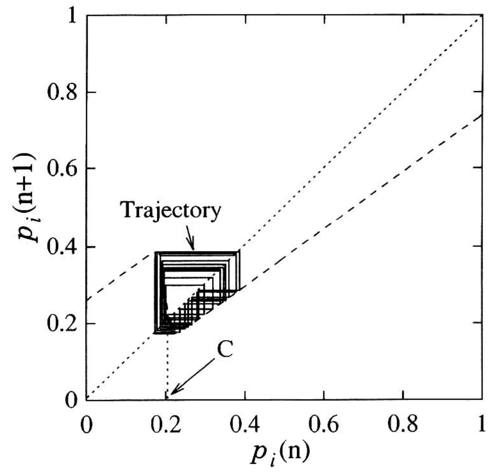

Fig. 2 The return map of the single chaotic neuron.

the point 'Z' is close to the point 'C'. At that time, it is clarified that the sign of $\overline{\Delta}_i$ is the same as the sign of Δ_i in all domain: $0 < p_i < 1$. Therefore these curves of Fig. 1 (b) and (d) have a bottom at the same position if β is small enough. Namely we can say that there is an attractor of the chaotic dynamics around the bottom of the quadratic function.

This consideration is also supported by a higher dimensional system.

Figure 3 shows the dynamics of the chaotic neural network made of two neurons ($M = 2$). This figure satisfies the following conditions: all eigenvalues of the weight matrix $W = [T_{ij}]$ are negative and the solution $\boldsymbol{p}^c$ of $W\boldsymbol{p}^c + \boldsymbol{I} = \boldsymbol{o}$ is in the Hypercube, where $\boldsymbol{I} = (I_1, \ldots, I_M)$ and the Hypercube is $[0,1]^M$. These conditions guarantee that $\boldsymbol{p}^c$ is a local minimum of the quadratic function; $F(\boldsymbol{p}) = -\frac{1}{2}\boldsymbol{p}^t W \boldsymbol{p} - \boldsymbol{I}^t \boldsymbol{p}$, and the local minimum exists in the Hypercube. $\boldsymbol{p}^c$ is illustrated in Fig. 3 as 'C', which corresponds to 'C' in Fig. 1. The line A and B correspond that the input for the neuron 1 and 2 are zero; $\sum_{j=1}^{M} T_{ij} p_j + I_i = 0$ $(i = 1, 2)$. From Eq. (9) and Fig. 1 (a), p_1 decreases in the right area of the line A and p_1 increases in the left area. Then p_2 decreases in the upper area of the line B and p_2 increases in the lower area. Assuming an initial state $\boldsymbol{p}(0)$ of Eqs. (8) and (9) is the point 'I', the state goes to the origin and comes close to the line A. When the state arrives to the line A or a neighborhood, in the next step it moves to the direction p_1 increasing (marked 'Jump' in Fig. 3). After the transition, the state will come close to the line B. In this way, it is proved theoretically that $\boldsymbol{p}(n)$ wanders around 'C'.

Figure 4 shows practical examples of four weight matrices which satisfy the above conditions. θ_1 and θ_2 are angles of these lines based on the p_2-axis and the p_1-axis respectively. $I_i/T = 0.067$, $\beta = 0.006$ and $\epsilon = 0.3$. From this figure, it is confirmed that the state wanders around of the local minimum.

As a result it is obvious that from the viewpoint of not $\boldsymbol{v}(n)$ but $\boldsymbol{p}(n)$, the dynamics is described by Eqs. (8) and (9) like a gradient descent. Then from the similarity of the sign of Δ_i and $\overline{\Delta}_i$, the dynamics is characterized from an asymptotically point of the linear system, that is a local minimum of a quadratic function. In other words, there is an attractor of the chaotic dynamics around a local minimum of a quadratic function. This result indicates that the chaotic neural network can roughly find out a local minimum point of a quadratic function using its attractor.

3. Linear Dynamical System and Objective Function

In the previous section we say that the chaotic neural network can roughly find out a local minimum point of a quadratic function using its attractor. However it is not clear that we can obtain a reasonable solution of a combinatorial optimization problem described as minimizing a quadratic function, that is, an objective function. The reason is that a local minimum point by the chaotic neural network exists inside the Hypercube. However the solution of a combinatorial optimization problem exists at a vertex of the Hypercube. We need a method of estimating a vertex from the obtained inside point. Additionally it must be guaranteed that the value of the quadratic function is small enough at the vertex.

In this article we adopt the step function as the above method. Concretely in case that an element of the inside point is positive, let the element be 1, and in otherwise, let the element be 0.

In this section we try to guarantee that the vertex estimated from the above method has a lower value of the objective function. First we define a linear dynam-

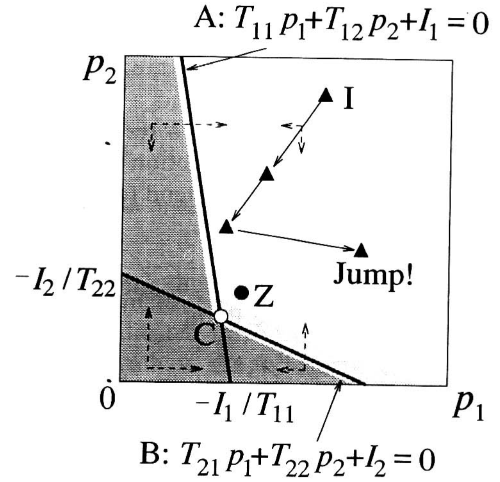

Fig. 3 Two dimensional state space of the chaotic neural network.

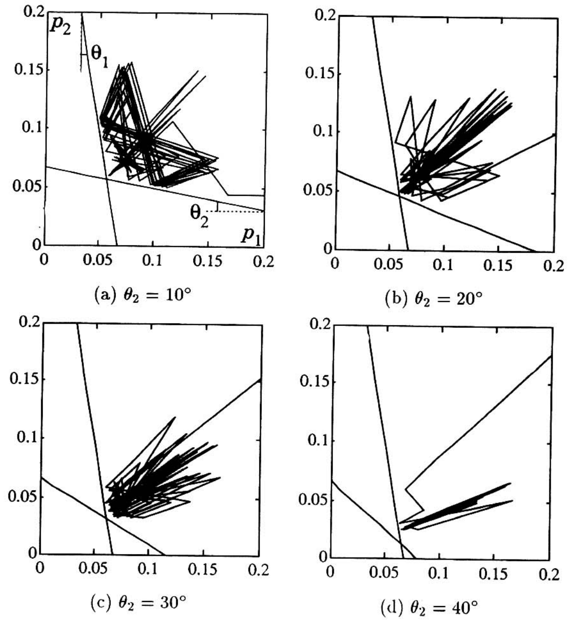

Fig. 4 Four attractors by the two dimensional system.

ical system and mention conditions for an asymptotically stable point of the system, corresponding to 'C' in Fig. 1. We next exactly investigate the relation between the asymptotically stable point and a value of the objective function at the vertex. As a result, we will have that a value of the objective function at a vertex with an asymptotically stable point is tend to be smaller than without an asymptotically one.

3.1 Definition of Linear Dynamical System and Asymptotically Stable Point

From the previous considerations, the state of the chaotic neural network wanders around the local minimum of the quadratic function. In this subsection, we define the local minimum as an asymptotically stable point of a linear dynamical system exactly.

First we define the linear dynamical system with a constraint. The system consists of variables x_i $(i = 1, 2, \ldots, M)$ which are constrained in a first quadrant. In this article the first quadrant is defined as $^\forall i; x_i \geqq 0$.

$$\frac{dx_i}{dt} = \begin{cases} 0, & (x_i = 0 \text{ and } (W\boldsymbol{x} + \boldsymbol{I})_i < 0) \\ (W\boldsymbol{x} + \boldsymbol{I})_i, & (\text{otherwise}) \end{cases} \tag{10}$$

where, $(W\boldsymbol{x}+\boldsymbol{I})_i$ is the ith element of the vector $W\boldsymbol{x}+\boldsymbol{I}$, that is $\sum_{j=1}^{M} T_{ij}x_j + I_i$, and $W = [T_{ij}]$ is an $M \times M$ symmetric matrix. This system is given by rewriting the difference equation with Δ_i in Sect. 2.2 to the differential equation. Then the point 'C' in Fig. 1 (c)(d) is an asymptotically stable point in Eq. (10).

We next define to expression of an asymptotically stable point. Let's assume that there is an asymptotically stable point $\boldsymbol{x}^*$ on a k dimensional coordinate plain, which is spread by k pieces of coordinate axes. Here the order of the element x_i^* $(i = 1, 2, \ldots, M)$ of the vector $\boldsymbol{x}^*$ is rearranged to $x_i^* > 0\,(i = 1, 2, \ldots, k)$ and $x_i^* = 0\,(i = k+1, \ldots, M)$, then we redefine the rearranged vector to $\boldsymbol{x}^*$ without loss of generality. The vector is separated to two vectors, $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$. The size of $\boldsymbol{x}_1^*$ is k and all elements are positive. The size of $\boldsymbol{x}_2^*$ is $M-k$ and all elements are zero. Additionally this rearrangement and separation of $\boldsymbol{x}^*$ are applied to the weight matrix W and the threshold vector $\boldsymbol{I}$;

$$W = \begin{pmatrix} W_1 & W_2^t \\ W_2 & W_3 \end{pmatrix} \text{ and } \boldsymbol{I} = (\boldsymbol{I}_1^t, \boldsymbol{I}_2^t)^t,$$

where, W_1, W_2 and W_3 are $k \times k$, $(M-k) \times k$ and $(M-k) \times (M-k)$ matrices respectively. $\boldsymbol{I}_1$ is a k dimensional vector. $\boldsymbol{I}_2$ is an $M-k$ dimensional vector. t represents transposition.

An example is shown by the following 3×3 matrix and the 3 dimensional vector;

$$W = \begin{pmatrix} -1 & -2 & -0.5 \\ -2 & -1 & -0.3 \\ -0.5 & -0.3 & -1 \end{pmatrix} \text{ and } \boldsymbol{I} = \begin{pmatrix} 0.05 \\ 0.05 \\ 0.05 \end{pmatrix}.$$

The linear dynamical system consisting of above W and $\boldsymbol{I}$ has an asymptotically stable point $(0, 0.038, 0.038)$ on the x_2-x_3 coordinate plain.

Considering the asymptotically stable point $\boldsymbol{x}^* = (0, 0.038, 0.038)$, (x_1, x_2, x_3) is transformed into (x_2, x_3, x_1) by rearrangement. Namely $\boldsymbol{x}^*$, W and $\boldsymbol{I}$ are rewritten to $(0.038, 0.038, 0)$,

$$W = \begin{pmatrix} -1 & -0.3 & -2 \\ -0.3 & -1 & -0.5 \\ -2 & -0.5 & -1 \end{pmatrix} \text{ and } \boldsymbol{I} = \begin{pmatrix} 0.05 \\ 0.05 \\ 0.05 \end{pmatrix}.$$

Note that $\boldsymbol{I}$ is not changed in this case. The plain 1, 2 and 3 before rearrangement are represented to plain 3, 1 and 2 respectively. The three plains $(W\boldsymbol{x}+\boldsymbol{I})_i = 0$ $(i = 1, 2, 3)$ are illustrated in Fig. 5 after rearrangement. In this figure ∘ is an asymptotically stable point.

Then separation is carried out as follows:

$$\boldsymbol{x}_1^* = (x_1^*, x_2^*) = (0.038, 0.038),\ \boldsymbol{x}_2^* = (x_3^*) = (0),$$

$$W_1 = \begin{pmatrix} -1 & -0.3 \\ -0.3 & -1 \end{pmatrix},\ W_2 = (-2 \quad -0.5),$$

$$W_3 = (-1),\ \boldsymbol{I}_1 = \begin{pmatrix} 0.05 \\ 0.05 \end{pmatrix} \text{ and } \boldsymbol{I}_2 = (0.05).$$

Secondly we turn our attention to the condition of $\boldsymbol{x}^*$ to be an asymptotically stable point. The conditions are expressed as follows.

[Conditions of $\boldsymbol{x}^*$ to be asymptotically stable]

1. $W_1\boldsymbol{x}_1^* + \boldsymbol{I}_1 = \mathbf{o}_1$ and $\boldsymbol{x}_2^* = \mathbf{o}_2$.
2. All eigenvalues of W_1 are negative.
3. $W_2\boldsymbol{x}_1^* + \boldsymbol{I}_2 < \mathbf{o}_2$.

where, $\mathbf{o}_1$ and $\mathbf{o}_2$ is k and $M-k$ dimensional zero vectors. $W\boldsymbol{x} + \boldsymbol{I} < \mathbf{o}$ means that all elements of the vector $W\boldsymbol{x} + \boldsymbol{I}$ are negative.

We say detail of the conditions and confirm the previous example.

Condition 1: The first condition is derived from that the right side of Eq. (10) is zero, On the example it is obvious from Fig. 5 that $\boldsymbol{x}^*$ exists on the x_1-x_2 coordinate plain, *i.e.* $x_3 = 0$, and on the plain 1 and 2 which

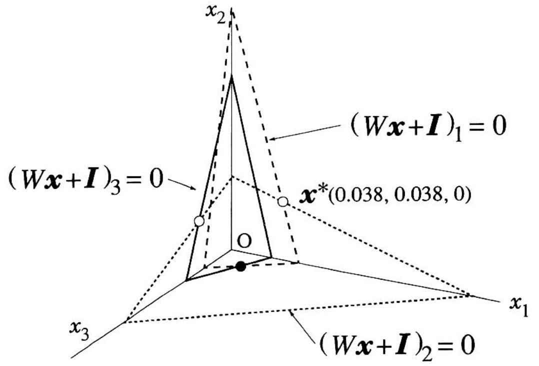

Fig. 5 The three plains and an asymptotically stable point of the linear dynamical system after rearrangement and separation.

correspond to $W_1\boldsymbol{x}_1^* + \boldsymbol{I}_1 = \mathbf{o}_1$. Hence the right side of Eq. (10) by $\boldsymbol{x}^*$ is zero, and namely $\boldsymbol{x}^*$ satisfies the first condition.

Condition 2: The second condition is derived from that $\boldsymbol{x}^*$ is an asymptotically stable point 'on the coordinate plain $\boldsymbol{x}_2 = \mathbf{o}_2$'. On the example $\boldsymbol{x}^*$ is on the coordinate plain $x_3 = 0$. All eigenvalues of W_1 are calculated as -1.3 and -0.7. It is well known on a linear dynamical system that all eigenvalues of a matrix being negative implies a equilibrium point being asymptotically stable. Namely $\boldsymbol{x}^*$ is an asymptotically stable point on $x_3 = 0$.

Condition 3: The third condition is derived from the inside of the parentheses in the upper line of the right side of Eq. (10). Combining the condition 1 and 3, $\boldsymbol{x}^*$ is guaranteed an equilibrium point. Because if $W_2\boldsymbol{x}_1^* + \boldsymbol{I}_2 > \mathbf{o}_2$ on the example, from Eq. (10) the state started from $\boldsymbol{x}^*$ can move to the direction x_3 increasing. In this case $\boldsymbol{x}^*$ is not asymptotically stable.

3.2 Objective Function and Asymptotically Stable Point

In this section we discuss relation of an asymptotically stable point and an objective function of a minimum searching problem.

First we define a minimum searching problem.

[**Minimum searching problem**]

In this article a minimum searching problem is expressed such as to find out the $\boldsymbol{y}$ at which the following quadratic function takes the global minimum.

$$F(\boldsymbol{y}) = -\frac{1}{2}\boldsymbol{y}^t A \boldsymbol{y} - \boldsymbol{b}^t \boldsymbol{y} \tag{11}$$

where, $\boldsymbol{y} = (y_1, y_2, \dots, y_M)^t, y_i = 1 \text{ or } 0 \, (i = 1, 2, \dots, M)$, $A = [-a_{ij}]$, $a_{ij} = a_{ji} \geqq 0$, $a_{ii} = 0$, $\boldsymbol{b} = (b_1, b_2, \dots, b_M)^t$, $b_i = b > 0$. This quadratic function is called the objective function. The objective function defined above is commonly used in a general combinatorial optimization problem, for example the traveling salesman problem and the N-Queen problem.

From the given objective function we can define the parameter of the linear dynamical system with a constraint as mentioned previous subsection.

$$W = [T_{ij}],\ T_{ij} = -a_{ij} \leqq 0\ (i \neq j),\ T_{ii} = -T < 0$$

$$\boldsymbol{I} = (I_1, I_2, \dots, I_M)^t,\ I_i = I = \gamma b > 0$$

Note that the values of T and γ (> 0) are decided independently of the given objective function.

We consider a vertex point $\boldsymbol{y}$ of the Hypercube which belongs to k dimensional coordinate plain. Here the order of the element y_i $(i = 1, 2, \dots, M)$ in $\boldsymbol{y}$ is rearranged to $y_i = 1\,(i = 1, 2, \dots, k)$ and $y_i = 0\,(i = k+1, \dots, M)$, then we redefine the rearranged vector to $\boldsymbol{y}$. The vector is separated to two vectors, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. The size of $\boldsymbol{y}_1$ is k and all elements are one. The size of $\boldsymbol{y}_2$ is $M - k$ and all elements are zero. Additionally rearrangement and separation of $\boldsymbol{x}^*$ are applied to the weight matrix W and the threshold vector $\boldsymbol{I}$ and $\boldsymbol{x}$.

$$W = \begin{pmatrix} W_1 W_2^t \\ W_2 W_3 \end{pmatrix},\quad \boldsymbol{I} = (\boldsymbol{I}_1^t, \boldsymbol{I}_2^t)^t \text{ and } \boldsymbol{x} = (\boldsymbol{x}_1^t, \boldsymbol{x}_2^t)^t.$$

On the example mentioned above, the vertex corresponding to $\boldsymbol{x}^* = (0.038, 0.038, 0)$ is $\boldsymbol{y} = (1, 1, 0)$ in Fig. 6. Then $\boldsymbol{y}_1 = (1, 1)$ and $\boldsymbol{y}_2 = (0)$. The value of the objective function at each vertex is calculated by Eq. (11) and indicated in Fig. 6.

Secondly we have the following lemmas easily.

Lemma 1: $F(\boldsymbol{y})$ is proportional to the sum of absolute value of all elements on ith row in W_1, which is represented as $|W_{1,i}| = \sum_{j=1}^{k} |T_{ij}|$.

Proof: Note $\boldsymbol{y}_2 = \mathbf{o}_2$, $T_{ij} < 0$ and $I_i = I$, $F(\boldsymbol{y})$ is calculated as follows:

$$F(\boldsymbol{y}) = -\frac{1}{2}\boldsymbol{y}^t W \boldsymbol{y} - \boldsymbol{I}^t \boldsymbol{y} = -\frac{1}{2}\boldsymbol{y}_1^t W_1 \boldsymbol{y}_1 - \boldsymbol{I}_1^t \boldsymbol{y}_1$$

$$= \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} |T_{ij}| - kI = \frac{1}{2}\sum_{i=1}^{k} |W_{1,i}| - kI.$$

□

We consider which the intersection $\overline{\boldsymbol{x}}$ of all plains $(W_1\boldsymbol{x}_1 + \boldsymbol{I}_1)_i = 0$ $(i = 1, 2, \dots, k)$ on the k dimensional coordinate plain is stable or not.

Lemma 2 The intersection $\overline{\boldsymbol{x}}$ of all plains $(W_1\boldsymbol{x}_1 + \boldsymbol{I}_1)_i = 0$ $(i = 1, 2, \dots, k)$ on the k dimensional coordinate plain is an asymptotically stable point if T and $|W_{2,i}|$ $(i = k+1, \dots, M)$ are large enough.

Proof: We consider the three conditions mentioned previous subsection.

It is clear that by its definition the intersection $\overline{\boldsymbol{x}}$ satisfies the first condition.

The second condition is satisfied if T is large enough, because it is well known that all eigen values of a matrix are negative if its diagonal elements are negative and large enough, from the Gerschgorin's theorem. In general T is defined independently of a given objective function. Therefore the second condition is satisfied for general objective functions if T is large enough.

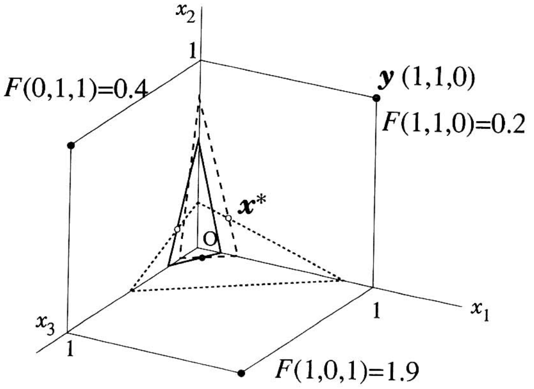

Fig. 6 The three vertices and values of the objective function.

Lastly it's the third condition. Let the minimum value in all elements of $\overline{\boldsymbol{x}}_1$ be $\overline{x}_{1,l}$. Note that all elements of $\overline{\boldsymbol{x}}_1$ are positive and all elements of W are negative, we have

$$(W_2\overline{\boldsymbol{x}}_1 + \boldsymbol{I}_2)_i < -\overline{x}_{1,l}|W_{2,i}| + I,$$

where $|W_{2,i}| = \sum_{j=1}^{k}|T_{ij}|$. Therefore in order to satisfy $(W_2\overline{\boldsymbol{x}}_1 + \boldsymbol{I}_2)_i < 0$ $(i = k+1, \ldots, M)$, $-\overline{x}_{1,l}|W_{2,i}| + I$ must be negative. In other words when $|W_{2,i}|$ is larger than $I/\overline{x}_{1,l}$, the intersection satisfies the third condition. □

From the lemma 1 and 2, we have the following theorem.

Theorem: There is an asymptotically stable point on a coordinate plain, if the objective function is small enough at the vertex $\boldsymbol{y}_1$ on the plain.

Proof: From the lemma 1, to find out a vertex $\boldsymbol{y}_1$ with a lower objective function means to decide of W_1 with small elements, and at the same time to decide of W_2^t with large elements from various separations of W. Note that the larger $|W_{2,i}^t|$ $(i = 1, \ldots, k)$ are, the larger $|W_{2,i}|$ $(i = k+1, \ldots, M)$ are, to find out a vertex with a lower objective function means to get larger $|W_{2,i}|$ $(i = k+1, \ldots, M)$. On the other hand from the lemma 2 if $|W_{2,i}|$ is large enough, there is an asymptotically stable point on the k dimensional coordinate plain defined by W_2. Namely, if the objective function is small enough at the vertex $\boldsymbol{y}_1$ on a coordinate plain, there is an asymptotically stable point on the plain. □

Lastly to confirm the theorem we have a simple numerical experiment. This experiment is carried out on the traveling salesman problem of 10 cities defined by Hopfield. We provide 254 vertices on Hypercube corresponding to a valid tour. Let each vertex be $\boldsymbol{y}$. Each $\boldsymbol{y}$ corresponds to a shorter tour and has a lower value of the objective function. From above rearrangement and separation, $\boldsymbol{y}$ is redefined and the coordinate plain corresponding to $\boldsymbol{y}_1$ is defined. In the coordinate plain we compute the intersection $\overline{\boldsymbol{x}}$ of all plains $(W_1\boldsymbol{x}_1 + \boldsymbol{I}_1)_i = 0$ $(i = 1, 2, \ldots, k)$ on the coordinate plain, which is a candidate for an asymptotically stable point. Here we investigate tour length of $\boldsymbol{y}$ and stability of $\overline{\boldsymbol{x}}$. Stability is estimated from the maximum value among $(W_2\overline{\boldsymbol{x}}_1 + \boldsymbol{I}_2)_i$ $(i = k+1, \ldots, M)$. Because when the value is negative, $W_2\overline{\boldsymbol{x}}_1 + \boldsymbol{I}_2 < \mathbf{o}_2$ can be satisfied and from the third condition we can say that $\overline{\boldsymbol{x}}$ is an asymptotically stable point.

The result is shown in Fig. 7. The horizontal axis shows a tour length of each $\boldsymbol{y}$. The vertical axis shows above stability. From Fig. 7 it is obvious that the stability of $\overline{\boldsymbol{x}}$ is roughly proportional to the tour length represented by $\boldsymbol{y}$ corresponding to $\overline{\boldsymbol{x}}$. Additionally $\overline{\boldsymbol{x}}$ with the global minimum is stable. This result supports that the objective function at the vertex on the coordinate plain with an asymptotically stable point is tend to be smaller than without an asymptotically one.

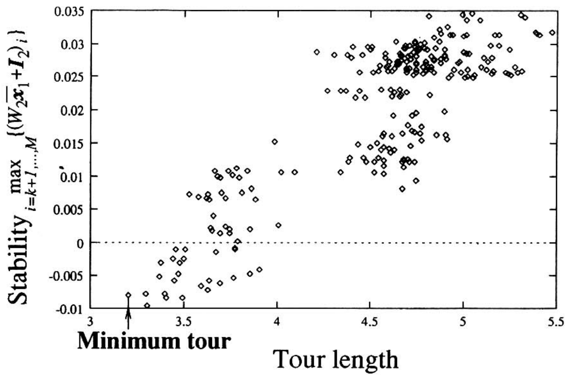

Fig. 7 Experimental result.

As a result, we can guarantee that the vertex derived from an asymptotically stable point has a lower value of the objective function. Combining the result of this subsection and previous section, it is guaranteed that we can obtain the reasonable solution of the minimum searching problem by the chaotic neural network if parameters T and γ are suitable.

3.3 Discussion

The theorem mentioned previous subsection is important for various analog recurrent neural network with negative self-feedback connections.

In the past study Uesaka [4] has analyzed stability of an analog recurrent neural network without self-feedback connections. It is very important for various minimum searching problem using this network. Because it exactly stated relation between an asymptotically stable point of an analog recurrent neural network without self-feedback and a local minimum of an objective function. We have analyzed an analog recurrent neural network with self-feedback connections [5]. It indicated to exist an equilibrium point on the coordinate plain. However it discussed to stability of a vertex and didn't say availability of the equilibrium point.

From the view point of self-feedback connections, it is considered that the theorem in this article is expansion of Uesaka's work. The theorem states relation between an asymptotically stable point of an analog recurrent neural network 'with negative self-feedback' and a vertex with a lower objective function. It is very important for various minimum searching problem using this network.

In this way the theorem is independent of the chaotic system. However the theorem is not effective for a convergence system, for example the Hopfield's model. In other words, The theorem doesn't say that we always get the global minimum using the Hopfield's network or linear system as mentioned above. Because a convergence system cannot escape a local minimum. From Fig. 7, there are several asymptotically stable points with not lower objective function. If the convergence system would be caught by these local minimums, we cannot get a suitable solution. On the other hand, the chaotic system has chaotic fluctuation (in this article this fluctuation is represented as a roughly gradient descent). Therefore using the fluctuation, the chaotic system can escape a local minimum.

4. Conclusion

We analyzed dynamics of the chaotic neural network and minimum searching principle of this network.

First the chaotic neural network was defined and its behavior was considered theoretically and experimentally. As a result we proved that the dynamics of the chaotic neural network is described like a gradient descent. Then we confirmed that the chaotic neural network can roughly find out a local minimum point of a quadratic function using its attractor. Secondly we guaranteed that the vertex corresponding a local minimum point derived from the chaotic neural network has a lower value of the objective function. Namely we resulted that the chaotic neural network can escape an invalid local minimum and find out a reasonable one.

In the future we must investigate property of escape from a local minimum by the chaos. From the viewpoint of the difference points between the chaos and non-chaos, we must discuss ability of the chaotic neural network.

References

[1] H. Nozawa, "A neural network model as a globally coupled map and applications based on chaos," Chaos, vol.2, no.3, pp.377–386, 1992.
[2] L. Chen and L. Aihara, "Transient Chaotic Neural Networks And Chaotic Simulated Annealing," Towards the Harnessing of Chaos, pp.347–352, Elsevier, 1994.
[3] J.J. Hopfield and D.W. Tank, " "Neural" computation of decisions in optimization problems," Biol. Cybern., vol.52, pp.141–152, 1985.
[4] Y. Uesaka, "Mathematical aspects of neuro-dynamics for combinatorial optimization," IEICE Trans., vol.E74, no.6, pp.1368–1372, 1991.
[5] M. Ohta, Y. Anzai, S. Yoneda and A. Ogihara, "A theoretical analysis of neural networks with nonzero diagonal elements," IEICE Trans., vol.E76–A, no.3, pp.284–291, 1993.

Masaya Ohta was born in Nara, Japan on August 24, 1965. He received the B.E. and M.E. degree in electrical engineering from University of Osaka Prefecture, Osaka, Japan in 1991 and 1993, respectively. He is currently working toward the Dr.Eng. degree in electrical engineering at the University of Osaka Prefecture. He is presently a graduate student at University of Osaka Prefecture. His research interests are in neural networks.

Kazumichi Matsumiya was born in Japan on January 15, 1972. He received the B.E. degree in electrical engineering from University of Osaka Prefecture, Osaka, Japan in 1995. He is currently working toward the Master Engineering degree in intelligence science at Tokyo Institute of Technology. He is presently a graduate student at Tokyo Institute of Technology. His research interests are in human visual mechanism.

Akio Ogihara was born in Japan on December 18, 1964. He received the B.E., M.E. and Dr.Eng. degrees in electrical engineering from University of Osaka Prefecture, Osaka, Japan, in 1987, 1989 and 1992, respectively. Since 1992 he has been with University of Osaka Prefecture, where he is now an assistant professor of College of Engineering. His current research interests include digital speech processing, neural networks, switched-capacitor filter and Fourier transform. Dr. Ogihara is a member of the Information Processing Society of Japan, the Acoustical Society of Japan and the IEEE.

Shinobu Takamatsu was born in Osaka Prefecture, on June 10, 1948. He received the B.E., the M.E., and the Ph.D. degrees in electrical engineering from the University of Osaka Prefecture in 1971, 1973 and 1978 respectively. He is an associate professor at the Department of Computer and Systems Sciences, the University of Osaka Prefecture. He is engaged in research concerning natural language processing and knowledge information processing. He is a member of Information Processing Society of Japan and Japanese Society for Artificial Intelligence.

Kunio Fukunaga was born in 1945 in Hyogo Prefecture, Japan. He received the B.E., M.E. and D.Eng. degrees in Electrical Engineering from University of Osaka Prefecture, Osaka, Japan 1967, 1969 and 1973 respectively. He has been with University of Osaka Prefecture since 1969, where he is now a professor in the Department of Computer and Systems Sciences. His current research interests include pattern recognition and image processing. He is a member of the Institute of Electronics, Information and Communication Engineers of Japan, the Information Processing Society of Japan, the Japanese Society for Artificial Intelligence and the IEEE.