Linear and Nonlinear Analysis
Yokohama Publishers
ISSN 2188-8167

Volume 10, Number 2, 2024, 111–121

ON CLASS OF NEARLY ABSORBING STATES IN SELF-ORGANIZING MAPS WITH INNER PRODUCT LEARNING MAPS

MITSUHIRO HOSHINO

Abstract. This paper deals in a mathematical framework with an iterative learning process using the inner product learning mapping in self-organizing maps known as the Kohonen type algorithm. On self-organizing maps, we discuss the mathematical properties of the model function, defined as a mapping from nodes to node values, in the iterative learning process. In particular, we focus on the absorbing and nearly absorbing properties, whereby a model function gradually takes on certain characteristics as learning progresses. In the inner product learning model, we give results on the local behavior and state preservation before and after learning, and a result on the existence of nearly absorbing states in the entire node array.

1. A mathematical model of self-organizing maps with inner product learning mapping

This paper deals with self-organizing maps, also referred to as the Kohonen type algorithm or the Kohonen Map [7], as an essential mathematical model and discusses it within a mathematical framework. This algorithm has the main functions of dimensionality reduction and clustering, and can be applied to statistics, with a wide range of applications, including cluster analysis, speech and text analysis, etc. It is also known for its application to combinatorial optimization, and is used to solve the traveling salesman problem; see, B. Angeniol [1].

On self-organizing maps, we discuss the mathematical properties of the model function, defined as a mapping from nodes to node values, in the iterative learning process. In particular, we focus on the absorbing and nearly absorbing properties, in which the state of the model function takes on certain characteristics as the learning process progresses. The mathematical state of model function allows us to evaluate whether the learning update is being performed appropriately. A linear combination type learning mapping is usually used for iterative learning in self-organizing maps, and is also used in many applications. The learning process using a linear combination type learning mapping has been discussed in a mathematical framework in several cases with essential and simple structures; see, for example, [2,

---

2020 *Mathematics Subject Classification.* 68T05.
Key words and phrases. Self-organizing maps, inner product learning, nearly absorbing state.
This work was partially supported by JSPS KAKENHI Grant Number JP23K03234 and the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto University.

3, 4, 5, 6, 8]. In this paper, we deal with the well-known dot product type and inner product type learning mappings, which are different in structure from the above, and discuss the mathematical properties related to the preservation of the state characteristics of the model function in the learning process. Dot product learning is a popular index for measuring similarity, and while there is little theoretical research or results from a mathematical approach to self-organizing maps, which uses dot product learning, it has long been used occasionally in applications.

In this paper, we deal with a self-organizing map with an inner product learning mapping, and assume a one-dimensional node array structure similar to the type introduced in T. Kohonen [7]. This model is characterized by four objects: *nodes*, *values of nodes*, *inputs*, and *learning process*, as follows:

$$\big(\{1,2,\dots,N\}, V, \{x_t\}_{t=0}^{\infty}, \{m_t(\cdot)\}_{t=0}^{\infty}\big).$$

(i) It is assumed that the model has nodes arranged according to a certain rule. Let I denote the set of all nodes, which is called the *node set.* In this paper, we suppose that the model consists of nodes aligned in a one-dimensional sequence and assume that I is a finite totally ordered set metrized by the following metric d, that is,

$$\begin{aligned} &I = \{1,2,\dots,N\} \subset \mathbb{N},\\ &1 < 2 < \cdots < N,\\ &d(i,j) = |i-j|, \quad i,j \in I. \end{aligned}$$

(ii) Each node has a *value* and it is updated by *learning.* V is the space of values of nodes. Suppose V is a subset of a real inner product space $(H, \langle\cdot,\cdot\rangle)$. A mapping $m : I \to V$ transforming each node i to its value $m(i)$ is called a *model function.*

(iii) X is the *input set.* Let X be a subset of V. Consider a sequence $\{x_t\}_{t=0}^{\infty} \subset X$ of input values.

(iv) The *learning process* is as follows. If an input is given, then each node learns the input value and updates its current value to a new value. Suppose an initial model function m_0 and a sequence $\{x_t\}_{t=0}^{\infty} \subset X$ of input values are given. Then the sequence $\{m_t(\cdot)\}_{t=0}^{\infty}$ of model functions are generated sequentially according to the following.

(a) Range of nodes whose values should be updated by learning: for each node $i \in I$,

$$N_1(i) = \big\{ j \in I \;\big|\; |j-i| \le 1 \big\}. \tag{1.1}$$

(b) Learning-rate factor: $\alpha > 0$.

(c) Learning process: for each m_t and x_t, let

$$M_t = \operatorname*{argmax}_{j \in I} \langle m_t(j), x_t \rangle$$

and define the sequence $\{m_t(\cdot)\}_{t=0}^{\infty}$ as follows.

$$(1.2)\qquad m_{t+1}(i) = \begin{cases} \dfrac{m_t(i)+\alpha x_t}{\|m_t(i)+\alpha x_t\|}, & \text{if } i \in \bigcup_{i^*\in M_t} N_1(i^*) \text{ and } m_t(i)+\alpha x_t \neq 0, \\ m_t(i), & \text{otherwise.} \end{cases}$$

It is also common to use $\cup_{i^*\in M_t} N_1(i^*)$ or $N_1(\min M_t)$ instead of $\{i \in \cup_{i^*\in M_t} N_1(i^*) \mid m_t(i)+\alpha x_t \neq 0\}$ as the set of nodes to be updated by learning. Here, as the number of update learning iterations increases, M_t becomes a singleton set with a high probability, so in the following discussion, there is no essential difference between $\cup_{i^*\in M_t} N_1(i^*)$ and $N_1(\min M_t)$. Furthermore, as learning progresses, except when the number of nodes is small, it becomes extremely unlikely that $m_t(i)+\alpha x_t = 0$ for some $i \in \cup_{i^*\in M_t} N_1(i^*)$ will hold. In this sense, we will use equation (1.2) in this paper.

It has been studied that the self-organizing map with a linear combination type learning mapping, that is, $m_{t+1}(i) = (1-\alpha_{m_t,x_t}(i))m_t(i) + \alpha_{m_t,x_t}(i)x_t$ for the rate $\alpha_{m_t,x_t}(i) \in [0,1]$, has some properties regarding the state preservation of model functions under learning processes. For example, the one-dimensional input model with a linear combination mapping has several characteristics, such as monotonicity being preserved before and after the model function is updated. A property like monotonicity is called an *absorbing state* or a *closed class of states* in a self-organizing map model in the sense that once a model function is in this state, it does not become any other state for any input. In addition, a state class in which the state is preserved for all nodes except for a few nodes before and after the model function is updated is called *nearly absorbing*. Furthermore, it has been shown that in some cases, nearly absorbing state classes exist; see, for example, [4]. Generally, when applying this type of algorithm to real data, problems arise during the learning update, such as whether the update is performed appropriately and whether the number of updates is sufficient. The above state characteristics help with the evaluation and judgment of such problems.

Numerical example 1. As an example of numerical calculation of the model in Section 1, consider a dot product self-organizing map with 150 nodes and 3-dimensional inputs, $\left(\{1,2,\ldots,150\}, V \subset \mathbb{R}^3, \{x_k\}_{k=0}^{\infty}, \{m_k(\cdot)\}_{k=0}^{\infty}\right)$. Initial node-values $m_0(i)$, $i = 1,2,\ldots,150$, are generated by the uniform distribution over $\{(x,y,z) \mid x^2+y^2+z^2 = 1\} \subset \mathbb{R}^3$. Inputs are generated by the uniform distribution over $[-1.05, 1.05]^3 \subset \mathbb{R}^3$. Assume Learning process (iv) in Section 1, where the dot product

$$\left\langle \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right\rangle = \sum_{i=1}^{3} a_i b_i, \quad \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbb{R}^3$$

is used as the inner product. In numerical calculation by computer, Figure 1 represents nodes and their values in each iteration steps. The position of every node means its value. It is observed that they are sorted by learning.

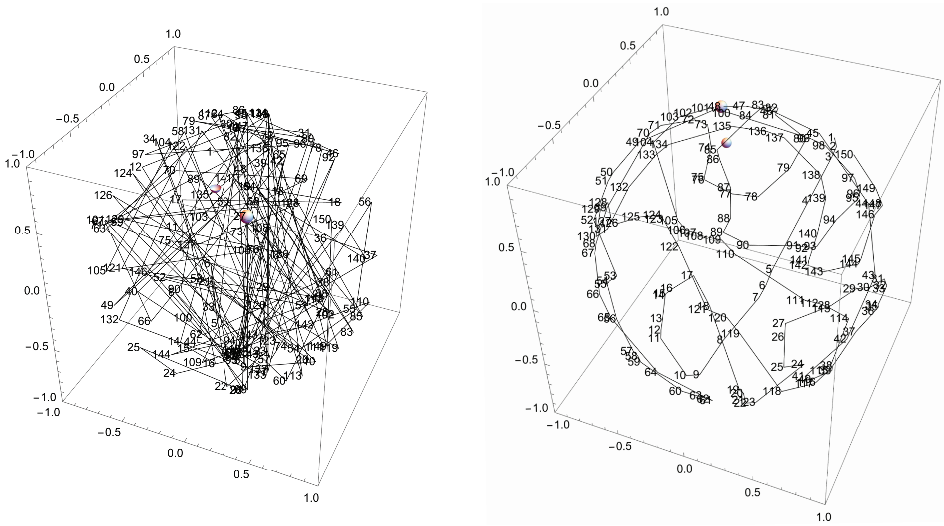

FIGURE 1. Initial values (left) and values after 5000 learning iterations (right) with $\alpha = 0.8$.

□

2. LOCAL BEHAVIOR AND STATE PRESERVATION OF UPDATED MODEL FUNCTIONS

In this section, in the model with inner product learning mapping, we will focus on the local behavior of the model function before and after learning. The following theorem describes its absorption properties and state preservation.

Theorem 2.1. *Suppose a self-organizing map model*

$$\big(\{1, 2, \ldots, N\}, V, \{x_t\}_{t=0}^{\infty}, \{m_t(\cdot)\}_{t=0}^{\infty}\big)$$

with an inner product learning mapping defined by (i)-(iv) in Section 1. Suppose that for the model function m after several updates, input x and three consecutive nodes $j-1, j, j+1$, $\operatorname{argmax}_{k\in I}\langle m(k), x\rangle$ *is a singleton set, and the following holds.*

$$\operatorname*{argmax}_{k\in I}\langle m(k), x\rangle = \{j\}, \tag{2.1}$$

$$\|m(j-1)\| = \|m(j)\| = \|m(j+1)\| = 1, \tag{2.2}$$

$$m(j \pm 1) + \alpha x \neq 0. \tag{2.3}$$

Let m' be the updated model function of m when learning input x. If the condition for inner product

$$\langle m(j-1) - m(j), m(j+1) - m(j)\rangle < 0 \tag{2.4}$$

holds for m and nodes $j-1$, j and $j+1$, then the inner product condition for m' and nodes $j-1$, j and $j+1$ also holds, that is,

$$\langle m'(j-1) - m'(j), m'(j+1) - m'(j)\rangle < 0. \tag{2.5}$$

Note that the statement of the above theorem still holds even if (2.4) and (2.5) are replaced with the following conditions

$$\langle m(j-1) - m(j), m(j+1) - m(j)\rangle \leq 0 \tag{2.6}$$

and

$$\langle m'(j-1) - m'(j), m'(j+1) - m'(j)\rangle \leq 0, \tag{2.7}$$

respectively, and this also holds for Theorem 2.2 and 2.3, which will be described below. Note further that in many situations, as the number of learning updates increases, the frequency and probability that $\operatorname{argmax}_{k\in I}\langle m(k), x\rangle$ becomes a singleton and all node values satisfy condition (2.2) gradually increase, and as the updates proceed further, these conditions are satisfied.

Proof. Using (2.2), we have

$$\begin{aligned}
&\|m(j) + \alpha x\|^2 - \|m(j-1) + \alpha x\|^2 \\
&\quad = \|m(j)\|^2 + 2\alpha\langle m(j), x\rangle + \alpha^2\|x\|^2 - \|m(j-1)\|^2 - 2\alpha\langle m(j-1), x\rangle - \alpha^2\|x\|^2 \\
&\quad = 2\alpha\Big(\langle m(j), x\rangle - \langle m(j-1), x\rangle\Big).
\end{aligned}$$

Assumption (2.1) implies $\langle m(j), x\rangle > \langle m(j-1), x\rangle$. Therefore, we obtain

$$\|m(j) + \alpha x\| > \|m(j-1) + \alpha x\|. \tag{2.8}$$

It follows from (2.3) and (2.8) that $m(i) + \alpha x \neq 0$ for $i = j, j \pm 1$. Therefore, by (iv) in Section 1 and equation (2.1) the following equation holds.

$$\begin{aligned}
&\langle m'(j-1) - m'(j), m'(j+1) - m'(j)\rangle \\
&\quad = \left\langle \frac{m(j-1) + \alpha x}{\|m(j-1) + \alpha x\|} - \frac{m(j) + \alpha x}{\|m(j) + \alpha x\|}, \frac{m(j+1) + \alpha x}{\|m(j+1) + \alpha x\|} - \frac{m(j) + \alpha x}{\|m(j) + \alpha x\|} \right\rangle \\
&\quad = A + B,
\end{aligned}$$

where

$$\begin{aligned}
A &= \frac{1}{\|m(j-1) + \alpha x\|\|m(j+1) + \alpha x\|}\langle m(j-1) - m(j), m(j+1) - m(j)\rangle, \\
B &= \frac{1}{\|m(j-1) + \alpha x\|}\left(\frac{1}{\|m(j+1) + \alpha x\|} - \frac{1}{\|m(j) + \alpha x\|}\right) \\
&\qquad\qquad \times \langle m(j-1) - m(j), m(j) + \alpha x\rangle \\
&\quad + \frac{1}{\|m(j+1) + \alpha x\|}\left(\frac{1}{\|m(j-1) + \alpha x\|} - \frac{1}{\|m(j) + \alpha x\|}\right) \\
&\qquad\qquad \times \langle m(j+1) - m(j), m(j) + \alpha x\rangle \\
&\quad + \left(\frac{1}{\|m(j-1) + \alpha x\|} - \frac{1}{\|m(j) + \alpha x\|}\right)\left(\frac{1}{\|m(j+1) + \alpha x\|} - \frac{1}{\|m(j) + \alpha x\|}\right) \\
&\qquad\qquad \times \|m(j) + \alpha x\|^2.
\end{aligned}$$

Inequality (2.4) implies $A < 0$. Furthermore, term B can be transformed into

$$\begin{aligned}
B = &\frac{1}{2\|m(j-1)+\alpha x\|}\left(\frac{1}{\|m(j+1)+\alpha x\|}-\frac{1}{\|m(j)+\alpha x\|}\right)\\
&\quad\times\big(\|m(j-1)+\alpha x\|^2-\|m(j-1)-m(j)\|^2-\|m(j)+\alpha x\|^2\big)\\
&+\frac{1}{2\|m(j+1)+\alpha x\|}\left(\frac{1}{\|m(j-1)+\alpha x\|}-\frac{1}{\|m(j)+\alpha x\|}\right)\\
&\quad\times\big(\|m(j+1)+\alpha x\|^2-\|m(j+1)-m(j)\|^2-\|m(j)+\alpha x\|^2\big)\\
&+\left(\frac{1}{\|m(j-1)+\alpha x\|}-\frac{1}{\|m(j)+\alpha x\|}\right)\left(\frac{1}{\|m(j+1)+\alpha x\|}-\frac{1}{\|m(j)+\alpha x\|}\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\|m(j)+\alpha x\|^2\\
=&-\frac{1}{2\|m(j-1)+\alpha x\|}\left(\frac{1}{\|m(j+1)+\alpha x\|}-\frac{1}{\|m(j)+\alpha x\|}\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\|m(j-1)-m(j)\|^2\\
&-\frac{1}{2\|m(j+1)+\alpha x\|}\left(\frac{1}{\|m(j-1)+\alpha x\|}-\frac{1}{\|m(j)+\alpha x\|}\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\|m(j+1)-m(j)\|^2\\
&-\frac{1}{2\|m(j-1)+\alpha x\|\|m(j)+\alpha x\|\|m(j+1)+\alpha x\|}\\
&\quad\times\big(\|m(j)+\alpha x\|-\|m(j-1)+\alpha x\|\big)\big(\|m(j)+\alpha x\|-\|m(j+1)+\alpha x\|\big)\\
&\quad\times\big(\|m(j-1)+\alpha x\|+\|m(j+1)+\alpha x\|\big).
\end{aligned}$$

In the same way as inequality (2.8) using (2.1) and (2.2),

$$\|m(j)+\alpha x\| > \|m(j+1)+\alpha x\|$$

can be shown. Using this inequality and (2.8), we obtain $B < 0$. Thus, inequality (2.5) holds for the updated model function m'. □

The following theorem is a result for the case where $\operatorname{argmax}_{k\in I}\langle m(k), x\rangle = \{j-1\}$ occurs instead of (2.1) in Theorem 2.1.

Theorem 2.2. *Suppose a self-organizing map model*

$$\big(\{1,2,\dots,N\}, V, \{x_k\}_{k=0}^{\infty}, \{m_k(\cdot)\}_{k=0}^{\infty}\big)$$

with an inner product learning mapping defined by (i)-(iv) in Section 1. Suppose that for the model function m after several updates, input x and three consecutive nodes $j-1, j, j+1$, $\operatorname{argmax}_{k\in I}\langle m(k), x\rangle$ is a singleton set, and the following holds.

$$\operatorname*{argmax}_{k\in I}\langle m(k), x\rangle = \{j-1\}, \tag{2.9}$$

$$\|m(j-1)\| = \|m(j)\| = \|m(j+1)\| = 1, \tag{2.10}$$

$$\|m(j)+\alpha x\| \geq 1, \tag{2.11}$$

$$\langle m(j-1)+\alpha x, m(j+1)\rangle \geq 0. \tag{2.12}$$

Let m' be the updated model function of m when learning input x. Then, the inner product condition for model function m and nodes $j-1$, j and $j+1$ is preserved for model function m' and nodes $j-1$, j and $j+1$, that is, if

$$\langle m(j-1) - m(j), m(j+1) - m(j)\rangle < 0, \tag{2.13}$$

then

$$\langle m'(j-1) - m'(j), m'(j+1) - m'(j)\rangle < 0. \tag{2.14}$$

Proof. By (2.9), $\langle m(j), x\rangle < \langle m(j-1), x\rangle$ holds and it follows from (2.10) that

$$\|m(j) + \alpha x\|^2 - \|m(j-1) + \alpha x\|^2 = 2\alpha\Big(\langle m(j), x\rangle - \langle m(j-1), x\rangle\Big) < 0.$$

Therefore, we have

$$\|m(j) + \alpha x\| < \|m(j-1) + \alpha x\|. \tag{2.15}$$

By (2.11) and (2.15), $m(j)+\alpha x \neq 0$ and $m(j-1)+\alpha x \neq 0$ hold. Using assumptions (2.9), (2.10), and (2.13), the following equation holds.

$$\begin{aligned}
&\big\langle m'(j-1) - m'(j),\ m'(j+1) - m'(j)\big\rangle \\
&\quad = \left\langle \frac{m(j-1)+\alpha x}{\|m(j-1)+\alpha x\|} - \frac{m(j)+\alpha x}{\|m(j)+\alpha x\|},\ m(j+1) - \frac{m(j)+\alpha x}{\|m(j)+\alpha x\|} \right\rangle \\
&\quad = \left\langle \frac{m(j-1)-m(j)}{\|m(j-1)+\alpha x\|},\ \frac{m(j+1)-m(j)}{\|m(j)+\alpha x\|} \right\rangle \\
&\qquad + \left\langle \frac{m(j-1)-m(j)}{\|m(j-1)+\alpha x\|},\ m(j+1) - \frac{m(j+1)+\alpha x}{\|m(j)+\alpha x\|} \right\rangle \\
&\qquad + \left(\frac{1}{\|m(j-1)+\alpha x\|} - \frac{1}{\|m(j)+\alpha x\|} \right) \\
&\qquad\quad \times \left\langle m(j)+\alpha x,\ m(j+1) - \frac{m(j)+\alpha x}{\|m(j)+\alpha x\|} \right\rangle.
\end{aligned}$$

We set

$$C = \left\langle \frac{m(j-1)-m(j)}{\|m(j-1)+\alpha x\|},\ \frac{m(j+1)-m(j)}{\|m(j)+\alpha x\|} \right\rangle,$$

then the previous equation can be transformed into the following equation using assumption (2.10).

$$\begin{aligned}
&\big\langle m'(j-1) - m'(j),\ m'(j+1) - m'(j)\big\rangle \\
&\quad = C + \frac{1}{\|m(j-1)+\alpha x\|}\left(1 - \frac{1}{\|m(j)+\alpha x\|}\right)\langle m(j-1)+\alpha x, m(j+1)\rangle \\
&\qquad - \frac{1}{\|m(j)+\alpha x\|}\left(1 - \frac{1}{\|m(j-1)+\alpha x\|}\right)\langle m(j)+\alpha x, m(j+1)\rangle \\
&\qquad - \frac{1}{\|m(j-1)+\alpha x\|\|m(j)+\alpha x\|}\big(\langle m(j-1)+\alpha x, \alpha x\rangle - \langle m(j)+\alpha x, \alpha x\rangle\big)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{\|m(j)+\alpha x\|}\left(\frac{1}{\|m(j-1)+\alpha x\|}-\frac{1}{\|m(j)+\alpha x\|}\right)\|m(j)+\alpha x\|^2 \\
&= C+D+E,
\end{aligned}
$$

where

$$
\begin{aligned}
D =& \frac{1}{\|m(j-1)+\alpha x\|}\left(1-\frac{1}{\|m(j)+\alpha x\|}\right) \\
&\qquad \times \big(\langle m(j-1), m(j+1)\rangle + \alpha\langle x, m(j+1)\rangle\big) \\
&- \frac{1}{\|m(j)+\alpha x\|}\left(1-\frac{1}{\|m(j-1)+\alpha x\|}\right) \\
&\qquad \times \big(\langle m(j), m(j+1)\rangle + \alpha\langle x, m(j+1)\rangle\big), \\
E =& -\frac{\|m(j-1)+\alpha x\|^2-\|m(j)+\alpha x\|^2}{2\|m(j-1)+\alpha x\|\|m(j)+\alpha x\|} - \frac{\|m(j)+\alpha x\|}{\|m(j-1)+\alpha x\|}+1.
\end{aligned}
$$

Here, inequality (2.13) implies $C < 0$. Since E can be transformed into

$$
E = -\frac{1}{2}\left(\frac{\sqrt{\|m(j)+\alpha x\|}}{\sqrt{\|m(j-1)+\alpha x\|}} - \frac{\sqrt{\|m(j-1)+\alpha x\|}}{\sqrt{\|m(j)+\alpha x\|}}\right)^2,
$$

we obtain $E < 0$. Furthermore, using (2.10) and (2.13), we have

$$
\langle m(j-1), m(j+1)\rangle < \langle m(j), m(j+1)\rangle. \tag{2.16}
$$

Using inequalities (2.11) and (2.15), we obtain

$$
\begin{aligned}
0 \leq& \frac{1}{\|m(j-1)+\alpha x\|}\left(1-\frac{1}{\|m(j)+\alpha x\|}\right) \\
&< \frac{1}{\|m(j)+\alpha x\|}\left(1-\frac{1}{\|m(j-1)+\alpha x\|}\right).
\end{aligned} \tag{2.17}
$$

Inequalities (2.12), (2.16), and (2.17) imply $D \leq 0$. Thus, it can be shown that inequality (2.14) holds. □

Now, in a similar way, we obtain the following theorem.

Theorem 2.3. *Suppose a self-organizing map model*

$$
\big(\{1,2,\dots,N\}, V, \{x_k\}_{k=0}^{\infty}, \{m_k(\cdot)\}_{k=0}^{\infty}\big)
$$

with an inner product learning mapping defined by (i)-(iv) in Section 1. Suppose that for the model function m after several updates, input x and three consecutive nodes $j-1, j, j+1$, $\operatorname{argmax}_{k\in I}\langle m(k), x\rangle$ is a singleton set, and the following holds.

$$
\operatorname*{argmax}_{k\in I}\langle m(k), x\rangle = \{j+1\}, \tag{2.18}
$$

$$
\|m(j-1)\| = \|m(j)\| = \|m(j+1)\| = 1, \tag{2.19}
$$

$$
\|m(j)+\alpha x\| \geq 1, \tag{2.20}
$$

$$
\langle m(j-1), m(j+1)+\alpha x\rangle \geq 0. \tag{2.21}
$$

Let m' be the updated model function of m when learning input x. Then, the inner product condition for m and nodes $j-1$, j and $j+1$ is preserved for m' and nodes $j-1$, j and $j+1$, that is, if

$$\langle m(j-1) - m(j), m(j+1) - m(j) \rangle < 0, \tag{2.22}$$

then

$$\langle m'(j-1) - m'(j), m'(j+1) - m'(j) \rangle < 0. \tag{2.23}$$

Proof. By using the symmetry of the node array, this theorem can be shown in the same way as Theorem 2.2. □

3. A nearly absorbing state class under inner product learning

By combining the results of Theorems 2.1, 2.2, and 2.3 as properties related to the state preservation of the model function before and after learning, it can be shown that under an inner product learning mapping, the state class of model functions that satisfy inequality (2.22) has the nearly absorbing property on the entire node array.

Theorem 3.1. *Suppose a self-organizing map model*

$$\big(\{1, 2, \ldots, N\}, V, \{x_k\}_{k=0}^{\infty}, \{m_k(\cdot)\}_{k=0}^{\infty}\big)$$

with an inner product learning mapping defined by (i)-(iv) in Section 1. Assume that

$$\|m(i)\| = 1, \quad i = 1, 2, \ldots, N, \tag{3.1}$$

$$\langle m(i), m(i+2) \rangle \geq 0, \quad i = 1, 2, \ldots, N-2 \tag{3.2}$$

for the model function m after several updates, and further assume that, for input x, the set $\mathrm{argmax}_{k \in I} \langle m(k), x \rangle$ is singleton, and

$$\langle m(i), x \rangle \geq 0, \quad i = j^* \pm 1, j^* \pm 2 \tag{3.3}$$

for the element $j^* \in \mathrm{argmax}_{k \in I} \langle m(k), x \rangle$. Let m' be the updated model function of m when learning input x. Then, the inner product condition is preserved for all i where $|i - j^*| \neq 2$, that is, if

$$\langle m(i-1) - m(i), m(i+1) - m(i) \rangle < 0, \tag{3.4}$$

then

$$\langle m'(i-1) - m'(i), m'(i+1) - m'(i) \rangle < 0. \tag{3.5}$$

Note that the statement of the above theorem holds even if the inequality signs in inequalities (3.4) and (3.5) are replaced with "$\leq$" and "$\leq$", respectively.

Proof. (i) When $i = j^*$, using assumptions (3.1) and (3.3), we have

$$\|m(j^* \pm 1) + \alpha x\|^2 = \|m(j^* \pm 1)\|^2 + 2\alpha\langle m(j^* \pm 1), x\rangle + \alpha^2\|x\|^2 \geq 1$$

and $m(j^* \pm 1) + \alpha x \neq 0$. Since $\text{argmax}_{k\in I}\langle m(k), x\rangle = \{i\}$, the inner product condition is preserved by Theorem 2.1.

(ii) When $i = j^* + 1$, $\text{argmax}_{k\in I}\langle m(k), x\rangle = \{i - 1\}$ holds. By assumptions (3.1) and (3.3),

$$\begin{aligned}\|m(i) + \alpha x\|^2 &= \|m(j^* + 1) + \alpha x\|^2 \\ &= \|m(j^* + 1)\|^2 + 2\alpha\langle m(j^* + 1), x\rangle + \alpha^2\|x\|^2 \geq 1\end{aligned}$$

holds, and we obtain $\|m(i) + \alpha x\| \geq 1$. Furthermore, by assumptions (3.2) and (3.3), we obtain

$$\begin{aligned}\langle m(i - 1) + \alpha x, m(i + 1)\rangle &= \langle m(j^*) + \alpha x, m(j^* + 2)\rangle \\ &= \langle m(j^*), m(j^* + 2)\rangle + \alpha\langle x, m(j^* + 2)\rangle \geq 0.\end{aligned}$$

According to these relations and Theorem 2.2, the inner product condition (3.4) is preserved.

(iii) When $i = j^* - 1$, $\text{argmax}_{k\in I}\langle m(k), x\rangle = \{i + 1\}$, so in the same way as when $i = j^* + 1$,

$$\|m(i) + \alpha x\|^2 = \|m(j^* - 1)\|^2 + 2\alpha\langle m(j^* - 1), x\rangle + \alpha^2\|x\|^2 \geq 1$$

holds from assumptions (3.1) and (3.3), so we obtain $\|m(i)+\alpha x\| \geq 1$. Furthermore, by assumptions (3.2) and (3.3), we have

$$\langle m(i - 1), m(i + 1) + \alpha x\rangle = \langle m(j^* - 2), m(j^*)\rangle + \alpha\langle m(j^* - 2), x\rangle \geq 0.$$

According to these relations and Theorem 2.3, the inner product condition (3.4) is preserved.

(iv) When $i = j^* \pm 3, j^* \pm 4, \ldots$, we have

$$\begin{aligned}m'(i - 1) &= m(i - 1), \\ m'(i) &= m(i), \\ m'(i + 1) &= m(i + 1).\end{aligned}$$

Therefore, the inner product condition (3.4) is preserved.

Thus, the theorem is proven from the above (i)-(iv). □

In the learning update, whether model function has come to satisfy the inner product condition (3.4) will help us evaluate and diagnose whether updates are being done properly and whether the number of updates is sufficient.

Acknowledgment

I wish to express my appreciation to the reviewers for their insightful comments on my paper.

References

[1] B. Angeniol, *Self-organizing feature maps and the travelling salesman problem*, Neural Networks **1** (1988), 289-293.

[2] M. Cottrell and J.-C. Fort, *Étude d'un processus d'auto-organisation*, Annales de l'Institut Henri Poincaré **23** (1987), 1–20 (in French).

[3] M. Cottrell, M. Olteanu, F. Rossi and N. Villa-Vialaneix, *Theoretical and applied aspects of the self-organizing maps*, in: Proceedings of the 11th international workshop WSOM 2016, pp. 3–26.

[4] M. Hoshino, *On states with absorbing tendencies in self-organizing maps with inputs in an inner product space*, Journal of Nonlinear Analysis and Optimization : Theory & Applications, **2** (2011), pp. 223 – 231

[5] M. Hoshino, *On state preserving property and nonexpansivity in self-organizing maps*, J. Nonlinear Convex Anal. **16** (2015), 2255–2261.

[6] M. Hoshino, *Local behavior of node with extreme value of model function in learning processes of basic self-organizing maps*, Linear and Nonlinear Anal. **6** (2020), 65–72.

[7] T. Kohonen, *Self-Organizing Maps, Third Edition*, Springer, 2001.

[8] T. Kohonen, *Essentials of the self-organizing map*, Neural Networks **37** (2013), 52–65.

Manuscript received 9 October 2024
revised 31 December 2024

Mitsuhiro Hoshino
Faculty of Systems Science and Technology, Akita Prefectural University
E-mail address: hoshino@akita-pu.ac.jp