LETTER — Communicated by James A. Bednar

Self-Organizing Maps with Asymmetric Neighborhood Function

Takaaki Aoki
aoki@acs.i.kyoto-u.ac.jp
Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, and CREST, JST, Kyoto 606-8501, Japan

Toshio Aoyagi
aoyagi@i.kyoto-u.ac.jp
Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

The self-organizing map (SOM) is an unsupervised learning method as well as a type of nonlinear principal component analysis that forms a topologically ordered mapping from the high-dimensional data space to a low-dimensional representation space. It has recently found wide applications in such areas as visualization, classification, and mining of various data. However, when the data sets to be processed are very large, a copious amount of time is often required to train the map, which seems to restrict the range of putative applications. One of the major culprits for this slow ordering time is that a kind of topological defect (e.g., a kink in one dimension or a twist in two dimensions) gets created in the map during training. Once such a defect appears in the map during training, the ordered map cannot be obtained until the defect is eliminated, for which the number of iterations required is typically several times larger than in the absence of the defect. In order to overcome this weakness, we propose that an asymmetric neighborhood function be used for the SOM algorithm. Compared with the commonly used symmetric neighborhood function, we found that an asymmetric neighborhood function accelerates the ordering process of the SOM algorithm, though this asymmetry tends to distort the generated ordered map. We demonstrate that the distortion of the map can be suppressed by improving the asymmetric neighborhood function SOM algorithm. The number of learning steps required for perfect ordering in the case of the one-dimensional SOM is numerically shown to be reduced from $O(N^3)$ to $O(N^2)$ with an asymmetric neighborhood function, even when the improved algorithm is used to get the final map without distortion.

1 Introduction

The self-organizing map (SOM) algorithm was proposed by Kohonen (1982, 2001) as a simplified neural network model having some essential properties to reproduce topographic representations observed in the brain (Hubel & Wiesel, 1962, 1974; von der Malsburg, 1973; Takeuchi & Amari, 1979). The SOM algorithm can be used to construct an ordered mapping from high-dimensional input data space onto low-dimensional array of units according to the topological relationships between input data. This implies that the SOM algorithm is capable of extracting the essential information from hugely complicated data. From the viewpoint of applied information processing, the SOM algorithm can be regarded as a generalized, nonlinear type of principal component analysis and has proven valuable in the fields of visualization, compression, and mining of various complicated data.

The learning performance becomes an important issue in many of the applications of the SOM algorithm. It is therefore desirable for real applications that fast convergence can be achieved to the map in which the correct topological order is produced. In addition, it is desirable for the resultant map to have as little distortion as possible to faithfully represent the structure of data. In this letter, we examine the effect of the form of the neighborhood function on the performance of the SOM learning algorithm. An inappropriate choice of neighborhood function may have a detrimental effect on learning performance. Erwin, Obermayer, and Schulten (1992) have reported that if the form of the symmetric neighborhood function is not convex, many undesirable metastable states will be present in the system. At each metastable state, the feature map of the SOM is ill structured, unlike from the correct map. Since the learning process becomes trapped in these metastable states, the system can escape from these undesirable states only through a sheer number of iterations because of their local stability. They have closely investigated certain relationships between the form of the symmetric neighborhood function and the existence of metastable states and have concluded that no metastable states exist when a convex neighborhood function is used in one-dimensional SOM. Hence, it is evident that a suitable selection of neighborhood function will pay dividends in the performance of the algorithm. However, even if such a suitable symmetrical neighborhood function is selected, there remains another important factor that spoils the learning performance. This is the emergence of a topological defect characterized by a globally conflicting point between multiple, locally ordered regions. A topological defect in the feature map appears occasionally during the learning process, especially when using a neighborhood function that is narrow compared with the total size of the SOM array. Figure 1A shows an example of the topological defect in a two-dimensional array of SOM with a uniform, rectangular input data space. The feature map is twisted at the center but topologically ordered everywhere else. Figure 1B shows an example of the topological defect in a one-dimensional array of SOM with scalar input data, which we

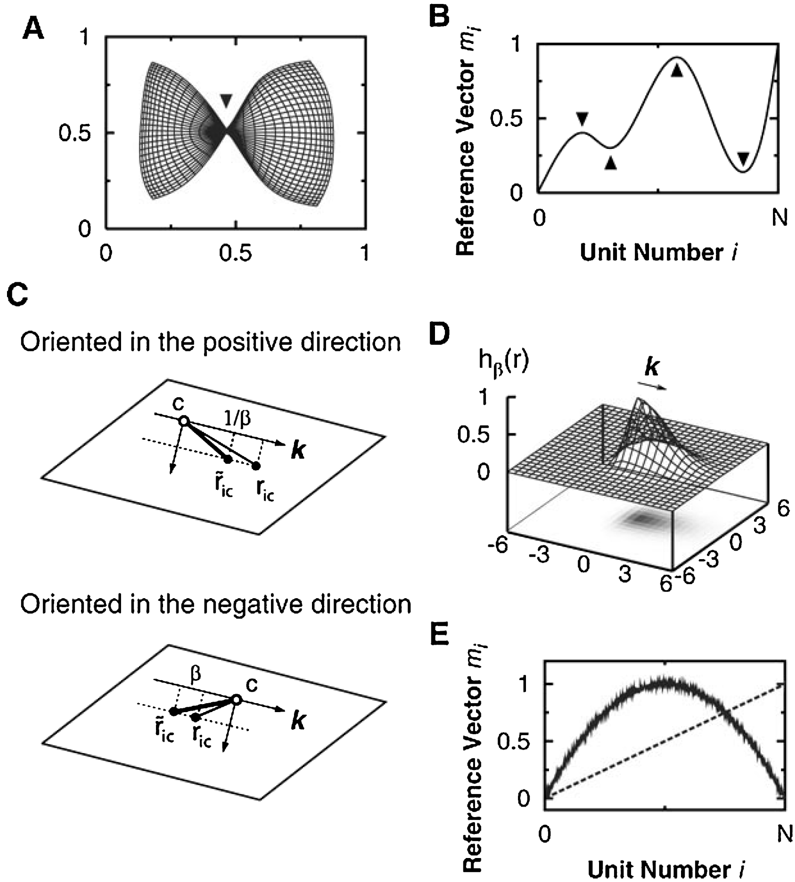


Figure 1: (A) Example of a topological defect in a two-dimensional array of SOM with a uniform rectangular input space. The triangle point indicates the conflicting point in the feature map. (B) Another example of topological defect in a one-dimensional array with scalar input data. The triangle points also indicate the conflicting points. (C) Method of generating an asymmetric neighborhood function by scaling the distance r_{ic} asymmetrically. The degree of asymmetry is parameterized by β. The distance of the node on the positive direction with asymmetric unit vector $\mathbf{k}$ is scaled by $1/\beta$. The distance on the negative direction is scaled by β. Therefore, the asymmetric gaussian function is described by $h_\beta(r_{ic}) = \frac{2}{\beta+1/\beta} h(\tilde{r}_{ic})$, where $\tilde{r}_{ic}$ is the scaled distance of node i from the winner c. (D) An example of an asymmetric gaussian function in two-dimensional SOM. (E) The initial reference vectors used in numerical simulations. There is a single kink point at the center of the array. In addition, the reference vectors are disturbed with small noise. The dashed line depicts one of the possible perfect maps, which should be formed from the input data given by the uniform distribution in [0,1].

call the *kink state* in this letter. In this kink state, the feature map is piecewise ordered on the scale of the width of neighborhood function, but there are one or more topological defects over the entire map. Once such a kink appears in the feature map, the learning process for getting the kink out of the map requires a large number of steps, which is typically several times larger than that without a kink (Geszti et al., 1990). This slow convergence time of learning due to a kink state is a serious problem that needs to be resolved, especially when large, complicated data are the inputs.

To avoid the trap in the kink state, several conventional and empirical methods have been used, such as finding a suitable choice of the initial reference vectors or rearranging the sequence of input vectors. In this letter, we demonstrate that the asymmetry of the neighborhood function can greatly improve the learning performance, particularly when a kink state appears during the learning process. The reason that the asymmetric neighborhood function is effective in the presence of the kink state is as follows.

In the process of getting a kink out of the feature map, the topological defect of the kink must move outside the boundary of the SOM array and vanish. Therefore, the convergence time of the learning process in the presence of the kink state is determined by the speed of the motion of topological defects. In general, the motive process of topological defects depends on the form of the neighborhood function. In the case of a symmetric gaussian function, which is commonly used in the conventional SOM algorithm, the movement of topological defects can be regarded as a random walk stochastic process, in which the motion of the defect behaves like a diffusion process. It is hypothesized that if an asymmetric neighborhood function is used, the motion of the defect will behave like a drift—a faster, more coherent process. Motivated by this idea, we investigate the effect of an asymmetric neighborhood function on the performance of the SOM algorithm, in particular, on the possibility that the asymmetry realizes faster-order learning than with a conventional symmetric neighborhood function in the presence of the topological defects.

2 Methods

2.1 One-Dimensional SOM. The SOM constructs a mapping from the input data space to the array of nodes that we call the feature map. To each node i, a parametric reference vector $\mathbf{m}_i$ is assigned. Through SOM learning, these reference vectors are rearranged according to the following iterative procedure. An input vector $\mathbf{x}(t)$ is presented at each time step t, and the best matching unit whose reference vector is closest to the given input vector $\mathbf{x}(t)$ is chosen. The distance between the input vector and the reference vector is here prescribed by the Euclidean distance $\|\mathbf{x}(t) - \mathbf{m}_i\|$ in the input data space. The best matching unit c, called the winner, is given by

$$c = \arg\min_i \|\mathbf{x}(t) - \mathbf{m}_i\|. \tag{2.1}$$

In other words, the data $\mathbf{x}(t)$ in the input data space are mapped on to the node c associated with the reference vector $\mathbf{m}_i$ closest to $\mathbf{x}(t)$. In SOM learning, the update rule for reference vectors is given by

$$\begin{aligned} \mathbf{m}_i(t+1) &= \mathbf{m}_i(t) + \alpha \cdot h(r_{ic})[\mathbf{x}(t) - \mathbf{m}_i(t)] \\ r_{ic} &\equiv \|\mathbf{r}_{ic}\| \equiv \|\mathbf{r}_i - \mathbf{r}_c\|, \end{aligned} \tag{2.2}$$

where α, the learning rate, is some small constant. The function $h(r)$ is called the neighborhood function, in which r_{ic} is the distance from the position $\mathbf{r}_c$ of the winner node c to the position $\mathbf{r}_i$ of a node i on the array of units. A widely used neighborhood function is the gaussian function defined by

$$h(r_{ic}) = \exp\left(-\frac{r_{ic}^2}{2\sigma^2}\right). \tag{2.3}$$

We expect an ordered mapping after iterating the above procedure a sufficient number of times.

In this letter, we investigate mainly the one-dimensional SOM, because its simplicity clarifies the essential dynamical process of SOM learning as a first step. Note that the reference vector in the case of the one-dimensional SOM is a scalar value (one-dimensional vector), In the following sections, we denote the reference vector by m_i and the input vector by x.

2.2 Asymmetric Neighborhood Function. The conventional neighborhood function widely used in practical applications is usually symmetric about the origin, that is, the position of the winner node; the gaussian function is an example of such a function. We now consider an asymmetric neighborhood function. We would like a method to transform any given symmetric neighborhood function to an asymmetric one and also to characterize the degree of asymmetry with a one-parameter variable. In addition, in order to single out the effect of asymmetry on the learning process, we require that the overall area of the neighborhood function, $\int_{-\infty}^{\infty} h(r)dr$, be preserved under the transformation from a symmetric to an asymmetric function. Keeping in the mind the above two points, the transformation of a given symmetric neighborhood function $h(r)$ proceeds as follows (see Figure 1C). Let us define an asymmetry parameter $\beta (\beta \geq 1)$, representing the degree of asymmetry and the unit vector $\mathbf{k}$ indicating the direction of asymmetry. If a unit i is located on the positive direction with $\mathbf{k}$, then the component parallel to $\mathbf{k}$ of the distance from the winner to the unit is scaled by $1/\beta$. If a unit i is located on the negative direction with $\mathbf{k}$, the parallel component of the distance is scaled by β. Hence, the asymmetric function

$h_\beta(r)$, transformed into its symmetric counterpart $h(r)$, is described by

$$h_\beta(r_{ic}) = A_\beta \cdot h(\tilde{r}_{ic}) \quad A_\beta = 2\left(\frac{1}{\beta} + \beta\right)^{-1}$$

$$\tilde{r}_{ic} = \begin{cases} \sqrt{\left(\frac{r_\parallel}{\beta}\right)^2 + \|\mathbf{r}_\perp\|^2}, & \text{if } \mathbf{r}_{ic} \cdot \mathbf{k} \geq 0 \\ \sqrt{\left(\beta r_\parallel\right)^2 + \|\mathbf{r}_\perp\|^2}, & \text{if } \mathbf{r}_{ic} \cdot \mathbf{k} < 0 \end{cases}, \tag{2.4}$$

where $\tilde{r}_{ic}$ is the scaled distance from the winner. $r_\parallel$ is the projected component of $\mathbf{r}_{ic}$, and $\mathbf{r}_\perp$ are the remaining components perpendicular to $\mathbf{k}$. For example, in 1D SOM, the positive direction of asymmetry is the increasing direction of the index of the SOM array. In this case, the asymmetric gaussian function is described by

$$h_\beta(r_{ic}) = \begin{cases} A_\beta \exp\left(-\frac{r_{ic}^2}{2\beta^2\sigma^2}\right), & \text{if } i \geq c \\ A_\beta \exp\left(-\frac{\beta^2 r_{ic}^2}{2\sigma^2}\right), & \text{if } i < c \end{cases}. \tag{2.5}$$

In the special case of the asymmetry parameter $\beta = 1$, $h_\beta(r)$ is equal to the original symmetric function $h(r)$. Figure 1D displays a example of asymmetric gaussian neighborhood functions in the two-dimensional array of SOM.

2.3 Topological Order and Distortion of the Feature Map. We define two measures that characterize the property of the feature map. One is the topological order η for quantifying the order of reference vectors in the SOM array. In a topologically ordered state of 1D SOM, the units of the SOM array should be arranged according to the magnitude of its reference vector m_i satisfying the condition

$$m_{i-1} \leq m_i \leq m_{i+1}, \quad \text{or} \quad m_{i-1} \geq m_i \geq m_{i+1}. \tag{2.6}$$

In a kink state, although a large population of the reference vectors satisfies the above order condition 2.6, there are topological defects that violate the order condition 2.6. On the feature map, we consider some domains within which the reference vectors satisfy the order condition 2.6. The topological order η can be defined as the ratio of the maximum domain size to the total number of units N, given by

$$\eta \equiv \frac{\max_l N_l}{N}, \tag{2.7}$$

where N_l is the size of domain l.

The other is the distortion χ, which measures the distortion of the feature map. The distribution of reference vectors obtained by SOM learning reproduces qualitative features of the probability density of the input vectors. There are several theoretical studies on the qualitative features of the distribution of the reference vectors depending on the probability density of input vectors (Ritter & Schulten, 1986; Ritter, 1991; Villmann & Claussen, 2006). However, as demonstrated in the following sections, the asymmetry of the neighborhood function tends to distort the distribution of reference vectors, which is quite different from the correct probability density of input vectors. For example, when the probability density of input vectors is uniform in the range [0, 1], a nonuniform distribution of reference vectors is formed through SOM learning with an asymmetric neighborhood function. Hence, for measuring the nonuniformity in the distribution of reference vectors, let us define the distortion χ. χ is a coefficient of variation of the size distribution of unit Voronoi tessellation cells and is given by

$$\chi = \frac{\sqrt{\mathrm{Var}(\Delta_i)}}{\mathrm{E}(\Delta_i)}, \tag{2.8}$$

where Δ_i is the size of Voronoi cell of unit i. To eliminate the boundary effect of the SOM algorithm, the Voronoi cells on the edges of the array are excluded from the calculation of the mean and variance of Δ_i. When the reference vectors are distributed uniformly, the distortion χ converges to 0.

2.4 Numerical Simulations. In the numerical simulations, we use the following parameter values: the total number of units $N = 1000$, the learning rate $\alpha = 0.05$ (constant), and the neighborhood radius $\sigma = 50$. The asymmetry parameter $\beta = 1.5$ and asymmetric direction $\mathbf{k}$ are set to the positive direction in the array. The aim is to examine the learning performance in the presence of a kink state; for this purpose, we use the initial condition that a single kink appears at the center of the array (see Figure 1E). The initial reference vectors are then given by

$$m_i = 4\frac{i(N-i)}{N^2} + \xi_i, \tag{2.9}$$

where ξ_i is a small noise that is uniform in the range -0.02 to 0.02. The probability density of input vectors $\{x\}$ is uniform distribution in [0, 1]. Because the density of input vectors is uniform, it follows that the desirable feature map is a linear arrangement of SOM nodes; an increasing map is shown as the dashed line in Figure 1E, and a decreasing one is not shown. In many numerical simulations, we confirmed the following result holds in a wide range of model parameters.

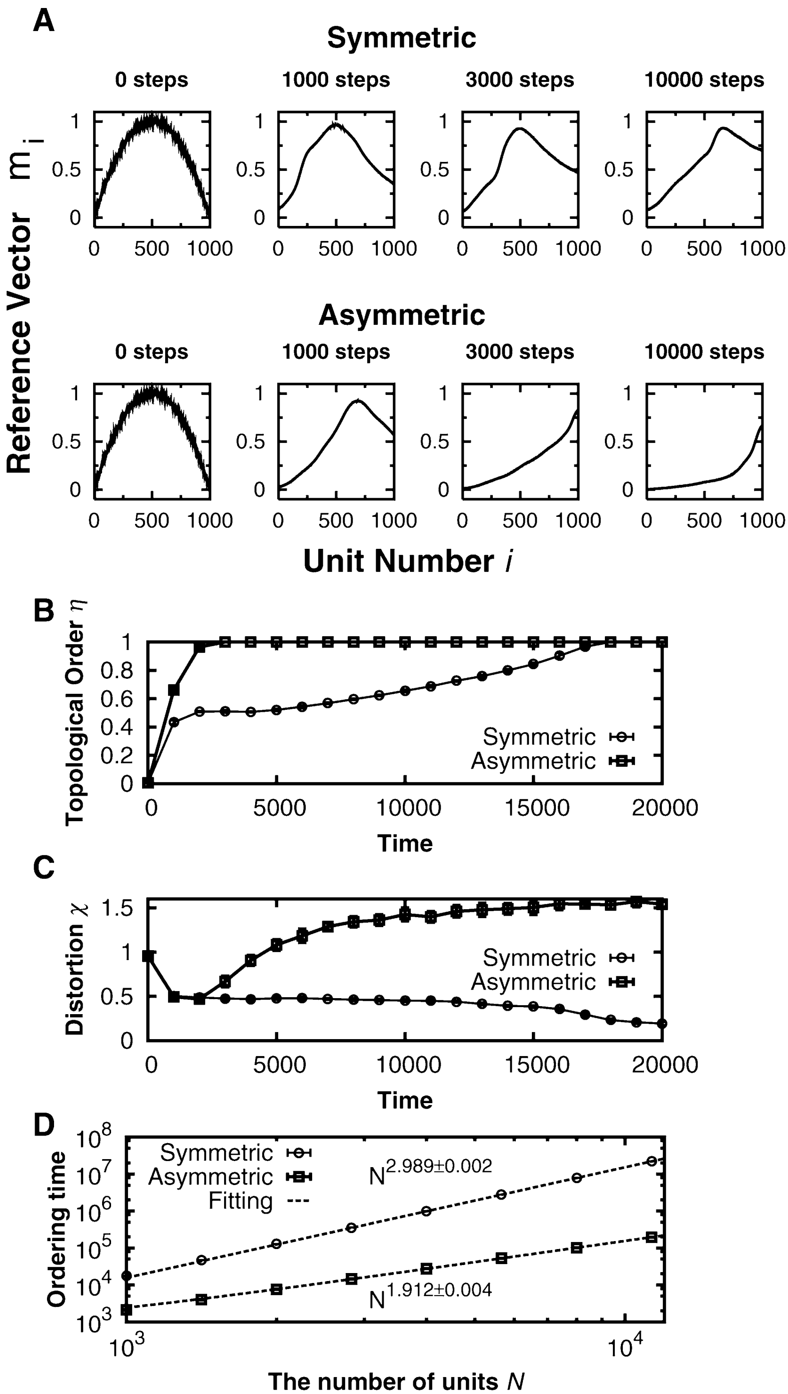
3 Results

In this section, we investigate the ordering process of SOM learning in the presence of a kink state in both symmetric and asymmetric cases of the neighborhood function. Here we use gaussian function for the original symmetric neighborhood function, as shown in equation 2.3.

Figure 2A shows a typical time development of the reference vectors m_i. In both cases, almost all topological defects due to the small noise ξ_i vanish within the first 1000 steps. However, in the case of the symmetric neighborhood function, a single kink point remains around the center of the array even after 10,000 steps. In contrast, in the case of the asymmetric function, this kink moves out of the feature map to the right so that the reference vectors are ordered within 3000 steps. This fast ordering process with an asymmetric neighborhood function can also be confirmed in Figure 2B, which demonstrates the time dependence of the topological order η. In the case of the asymmetric neighborhood function, η rapidly converges to 1 (completely ordered state) within 3000 steps. For the symmetric function, η also increases up to 0.5 within 3000 steps, which results from the elimination of almost all the topological defects made by the initial small noise. However, the process of eliminating the last single kink takes a large amount of time (about 18,000 steps). To quantify the performance of the ordering process, let us define the ordering time as the time at which η reaches 1 (completely ordered state). Figure 2D shows the ordering time as a function of the total number of units N for both asymmetric and symmetric cases of the neighborhood function. It is found that the ordering time scales roughly as N^3 and N^2 for symmetric and asymmetric neighborhood functions, respectively. Therefore, the asymmetric neighborhood function accelerates the learning process in the presence of a kink state, with performance being improved by about one order. We should remark that the range σ of the neighborhood function in Figure 2D is fixed when N is changed. With a fixed σ, the increase of the units N can improve the

---

Figure 2: The asymmetric neighborhood function enhances the ordering process of SOM, though this asymmetry causes the distortion of the feature map. (A) A typical time development of the reference vectors m_i in cases of symmetric and asymmetric neighborhood functions. The input vectors are randomly selected from a uniform distribution in [0, 1]. (B) Time dependence of the topological order η for symmetric and asymmetric neighborhood functions, estimated from 10 trials. The standard deviations are denoted by the error bars, which cannot be seen because they are smaller than the size of the graphed symbol. (C) Time dependence of the distortion χ. (D) Ordering time as a function of the total number of units N. Each data point is estimated from 12 trials. The fitting function is described by $Const. \cdot N^{\gamma}$.


A
Symmetric
0 steps
1000 steps
3000 steps
10000 steps
Asymmetric
0 steps
1000 steps
3000 steps
10000 steps
Reference Vector m_i
Unit Number i
B
Topological Order η
Symmetric
Asymmetric
Time
C
Distortion χ
Symmetric
Asymmetric
Time
D
Ordering time
Symmetric
Asymmetric
Fitting
N^2.989±0.002
N^1.912±0.004
The number of units N

resolution of the representation in the input space. However, more learning steps are required for achieving the ordered state. We will also discuss the case that σ is scaled proportional to N in discussion.

However, one problem arises in the feature map obtained through the learning process with the asymmetric neighborhood function. After 10,000 steps, the distribution of the reference vectors in the feature map develops an unusual bias (see Figure 2A). In general, the distribution of reference vectors in the learned feature map should represent the probability density of input vectors, which is uniform in [0, 1]. In the case of the symmetric neighborhood function, the magnification law of the one-dimensional SOM has been studied theoretically (Ritter & Schulten, 1986). By this magnification law, the uniform input vectors create a uniformly linear map. However, this result cannot be applied to the case of the asymmetric neighborhood function. Furthermore, the nonuniform, distorted feature map is formed by the drift force of the asymmetric neighborhood function. To measure this distortion of the feature map, we adopt the distortion χ (see equation 2.8), which is the coefficient of variation of the size distribution of Voronoi cells. Figure 2C shows the time dependence of the distortion χ during learning. In the case of the symmetric neighborhood function, χ eventually converges to almost 0. This result indicates that the feature map obtained with the symmetric neighborhood function has an almost uniform size distribution of Voronoi cells, which is the result of the uniform probability density of input vectors. In contrast, in the case of the asymmetric function, χ converges to a finite value ($\neq 0$).

To solve this distorted feature map problem, we introduce an improved algorithm for the asymmetric neighborhood function. The improved algorithm includes two novel steps during learning. First, we introduce an inversion operation on the direction of the asymmetric neighborhood function. After every time interval T, the direction of the asymmetry is turned in the opposite direction, as illustrated in Figure 3. Since the distortion in the feature map is along the direction of the asymmetry of the neighborhood function, this periodic inversion is expected to average out the distortion in the feature map, leading to the formation of a feature map without distortion. We set $T = 3000$ in the following numerical simulations. It is noted that the interval T should be set to a larger value than the typical ordering time for the asymmetric neighborhood function. The optimal value of the flip period T will be considered in the discussion. Next, we introduce an operation that decreases the degree of asymmetry of the neighborhood function. The degree of asymmetry can be controlled by one parameter, β. When $\beta = 1$, the neighborhood function equals the original symmetric function. With this operation, β is decreased to 1 with each time step, as illustrated in Figure 3. In our numerical simulations, we adopt a linear decreasing function,

$$\beta = 1 + (\beta_0 - 1)\left(1 - \frac{t}{t_{\mathrm{Total}}}\right), \tag{3.1}$$

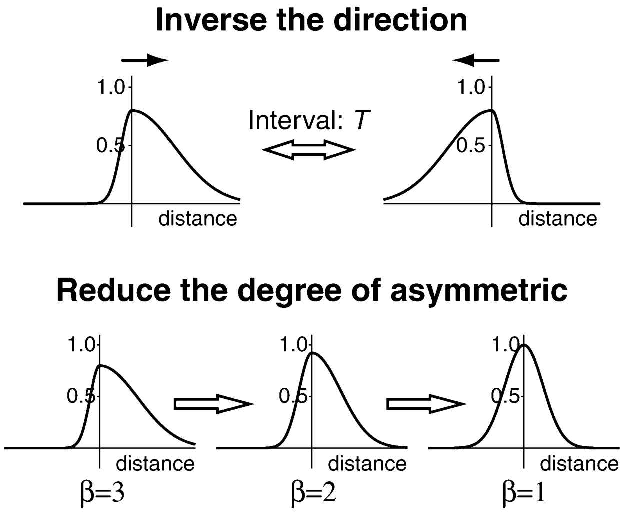


Figure 3: Illustration of the improved algorithm for asymmetric neighborhood function.

where t_{Total} is the total time step and β_0 is the initial constant value of the asymmetry parameter β.

Using this improved algorithm for the asymmetric neighborhood function, the results obtained are summarized in Figure 4. It is observed that the improved algorithm preserves the faster-order learning (see Figure 4A). Furthermore, as shown in Figure 4C, the ordering time scales as N^2 for the improved algorithm, which is the same as the original asymmetric neighborhood function. Therefore, the improved algorithm has the ability to accelerate the learning process. This improved algorithm also suppresses the distortion of the feature map. As shown in Figure 4B, the distortion χ converges to almost 0, which implies that the feature map without distortion represents the probability density of input vectors. This result can also be confirmed by another measure of distortion on the feature map. Der, Herrmann, and Villmann (1997) proposed an effective visualization method for the topological defects and distortions on the feature map by using the spatial spectral density of the deviation from uniform ordered feature map. When the deviation u_i is defined by $u_i \equiv m_i - \frac{1}{2}(m_{i-1} + m_{i+1})$, the spectral density $S(k)$ is given by

$$S(k) = \|\tilde{u}_k\|^2, \quad \tilde{u}_k = \frac{1}{N}\sum_{n} \exp\left(\frac{2\pi ikn}{N}\right) u_n. \tag{3.2}$$

When the topological defects or distortions exist on the feature map, the characteristic peaks in the spectral density appear at the wave number corresponding to the typical spatial scale of the structures in the feature map. When the feature map is formed successfully, the spectral density becomes nearly flat and has no peaks. Figure 4D shows the time development of the spectral density for the cases of the symmetric, the asymmetric, and the improved asymmetric neighborhood functions. We can see that for the improved asymmetric neighborhood function, the spectral density rapidly converges to a flat distribution without peaks, whereas some significant peaks still remain for the original asymmetric neighborhood function. Although the density finally converges to a flat distribution in the case of the symmetric neighborhood function, the speed of the convergence is much slower than that for the improved asymmetric neighborhood function.

In summary, by using the improved algorithm of asymmetric neighborhood function, we confer the full benefit of both the fast-order learning and the undistorted feature map.

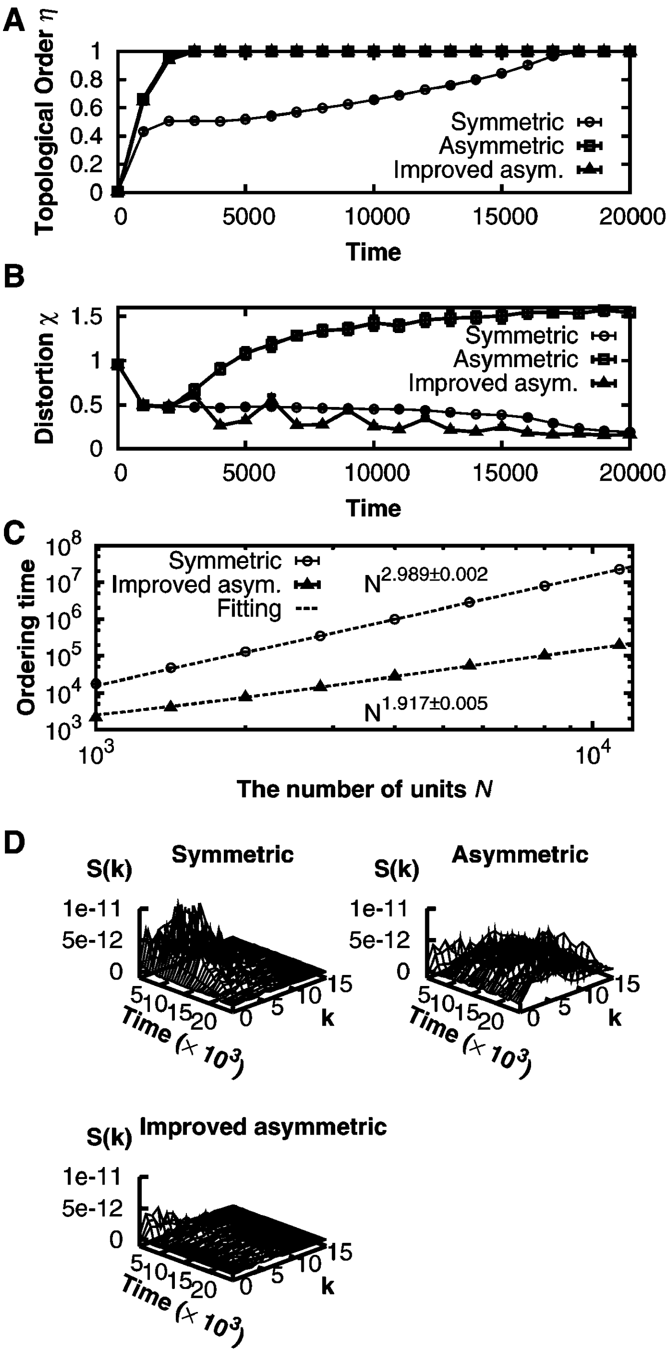
4 Applicability to More General Situations

In this section, we demonstrate that the asymmetric neighborhood function retains its useful properties in more general situations, such as the one-dimensional ring SOM (periodic boundary condition on the SOM lattice) and the two-dimensional SOM. These results suggest that the proposed method can be widely applied to various types of real data.

4.1 One-Dimensional Ring SOM. Let us consider the case of the ring of SOM units, which is useful when one-dimensional input data are periodic. For example, we assume a uniform distribution on the unit circle as input data, which is then given by

$$\mathbf{x}(t) = \{\cos\theta(t), \sin\theta(t)\}, \qquad \theta(t) \text{ is uniform in } [0, 2\pi)\,. \tag{4.1}$$

Figure 4: The improved algorithm utilizing an asymmetric neighborhood function has the ability not only to expedite the learning process but also to suppress the distortion of the feature map. (A) Time dependence of the topological order η (interval $T = 3000$). After 3000 steps, the SOM array is perfectly ordered in the case of the improved algorithm for asymmetric neighborhood function, as in the asymmetric neighborhood function. Therefore, these data points are overlapping at $\eta = 1$. Note that the error bars cannot be seen owing to their smallness, as in Figure 2B. (B) Time dependence of the distortion χ, in the same situation as *A*. (C) Ordering time as a function of the total number of units *N*. (D) Another measure of the distortion on the feature map. The spectral density $S(k)$ is the Fourier components of local deformation of the feature map. A salient peak in the spectral density indicates the existence of the distortion or topological defect in the feature map.

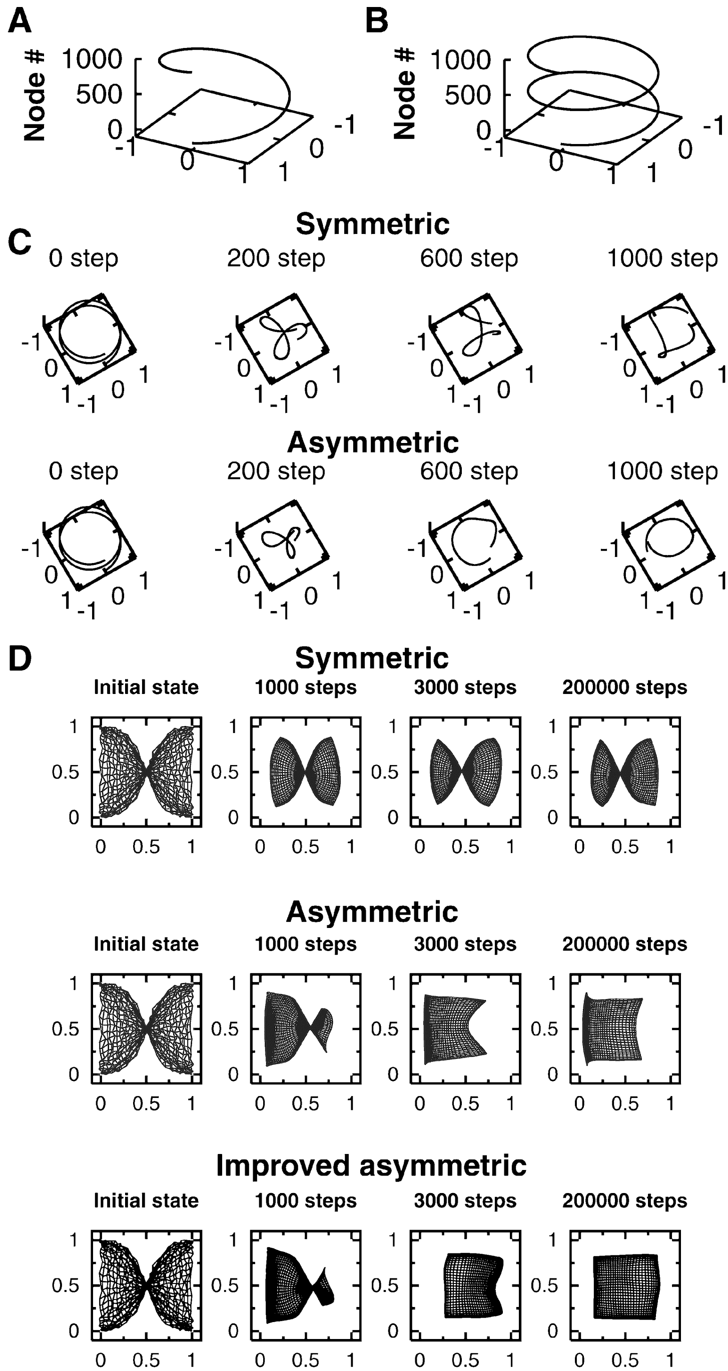

A
Topological Order η
Symmetric
Asymmetric
Improved asym.
Time
B
Distortion χ
Symmetric
Asymmetric
Improved asym.
Time
C
Ordering time
Symmetric
Improved asym.
Fitting
$N^{2.989\pm0.002}$
$N^{1.917\pm0.005}$
The number of units N
D
S(k)
Symmetric
S(k)
Asymmetric
Time (× 10³)
k
S(k)
Improved asymmetric
Time (× 10³)
k

For such input data, it is desirable that the position of the SOM units makes one rotation as θ varies over one period (see Figure 5A). To start in an artificial kink-like state, we set the initial reference vectors so that the position of the units makes two rotations over one period of θ $[0,2\pi]$, as shown in Figure 5B. The model parameters used in these simulations are the same as in the previous section. Figure 5C shows a typical time development of the reference vectors. During learning, the system should first destroy the initial nonoptimal ordered state and rearrange the ordered state by some intermediate disordered state. We can see from Figure 5C that the asymmetric neighborhood function can achieve faster ordering than the symmetric one. Hence, one can say that the asymmetry of the neighborhood function is also effective for quick reformation from the kink-like undesirable state to the optimal state through the intermediate disordered state.

4.2 Two-Dimensional SOM. Next, we show some results from the two-dimensional SOM, which is also important in practical applications. We demonstrate that a similarly fast ordering process can be realized with an improved asymmetric neighborhood function in two-dimensional SOM (see Figure 5D). The conventional symmetric neighborhood function has trouble in correcting the twisted state in the two-dimensional SOM. However, using the improved asymmetric neighborhood function, the feature map converges to the completely ordered map in much less time, whereas for the case of the symmetric neighborhood function, this twisted state cannot be corrected even in 200,000 steps. This suggests that the asymmetric neighborhood function is also effective in improving the convergence process in higher-dimensional SOMs.

Finally, although the advantage of the asymmetric neighborhood function is also seen for the above two cases, further study is required to examine this improvement of convergence time with the asymmetric neighborhood function.

Figure 5: (A) Ordered state of one-dimensional ring SOM lattice for the given input data, $\mathbf{x}(t) = \{\cos\theta(t), \sin\theta(t)\}$, where $\theta(t)$ is randomly generated according to a uniform distribution in $[0, 2\pi)$. (B) Initial reference vectors intended to create a kink-like state during the learning process. (C) Comparison of typical ordering behaviors in the 1D ring SOM between the symmetric and the asymmetric neighborhood functions. (D) Typical time development of reference vectors in a two-dimensional array of SOM for the cases of symmetric, asymmetric, and improved asymmetric neighborhood functions. The input vectors are randomly selected from a uniform distribution $[0, 1] \times [0, 1]$. Simulation parameters: $\alpha = 0.05$, $\beta = 1.5$, $\sigma = 5$, $N = 30 \times 30$, and $T = 2000$.


A
Node #
1000
500
0
-1
0
1
-1
0
1
B
Node #
1000
500
0
C
Symmetric
0 step
200 step
600 step
1000 step
Asymmetric
0 step
200 step
600 step
1000 step
D
Symmetric
Initial state
1000 steps
3000 steps
200000 steps
Asymmetric
Initial state
1000 steps
3000 steps
200000 steps
Improved asymmetric
Initial state
1000 steps
3000 steps
200000 steps
0.5

5 Discussion

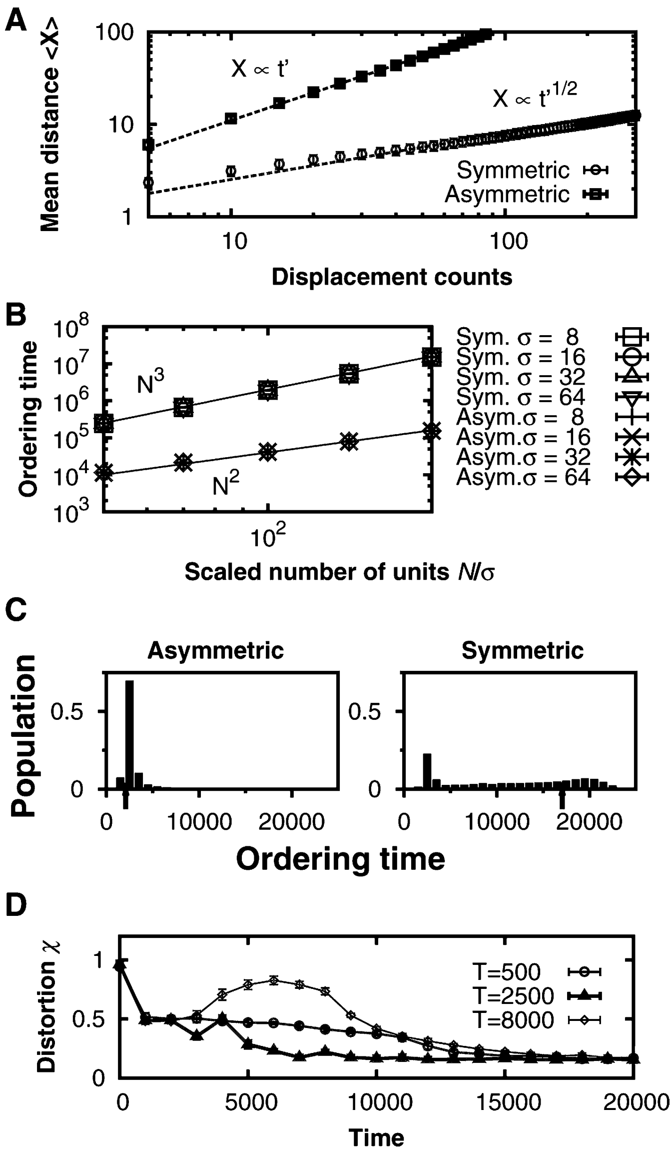
In this letter, we discussed the learning process of the self-organized map, especially in the presence of a kink (1D topological defect). Once a kink in the feature map appears, the learning process tends to take a very long time to eliminate this kink in the map and achieve the complete ordered state. Interestingly, even in the presence of the kink, we found that the asymmetry of the neighborhood function enables the system to accelerate the learning process. Compared with the conventional symmetric function with a constant neighborhood range, the convergence time of the learning process can be roughly reduced from $O(N^3)$ to $O(N^2)$ (N is the total number of units).

The difference in the scaling exponent of the ordering time can be explained as follows. In order to produce the correct topological order, all topological defects need to move out of the feature map and vanish, so the process of movement of the topological detect determines the performance of the ordering process. Let us define $X(\hat{t})$ as the net distance from the initial position of the topological defect, as a function of the actual steps $\hat{t}$ in which the defect moves, where $\hat{t}$ is the number of times the defect has actually stepped in t learning steps. Figure 6A shows the mean distance $\langle X(\hat{t})\rangle$ as a function of the actual steps $\hat{t}$ the defect has moved, estimated from 1000 samples ($\alpha = 0.05$, $\beta = 1.5$, $\sigma = 50$, and $N = 1000$). For the case of a symmetric neighborhood function, the mean distance obeys

$$\langle X(\hat{t})\rangle \propto \hat{t}^{1/2}. \tag{5.1}$$

This result indicates that the movement of the topological defect is well described by a random walk because of the equal probability of defect

Figure 6: (A) Average distance $\langle X(\hat{t})\rangle$ of the topological defect from the initial point as a function of the actual steps $\hat{t}$ in which the defect moved. $\hat{t}$ is the number of times the defect hat actually stepped in t steps. Each data point is estimated from 1000 simulations. (B) Effect of the width of neighborhood function. Log-log plot of ordering time versus scaled number of units N/σ, with $\sigma = 8,16,32,64$. The fitting functions are $Const. \cdot (N/\sigma)^k, \quad k = 2, 3$. Each data point is estimated from 12 simulations. (C) Distribution of ordering times when the initial reference vectors are generated by a uniform random value in [0, 1], estimated from 10,000 simulations. The arrow indicates that the mean ordering time in the case of the single kink initial condition. (D) The optimal value of the flip period T in the improved asymmetric neighborhood function. Three typical cases are presented: short ($T = 500$), around optimal ($T = 2500$), and long time ($T = 8000$).


A
Mean distance <X>
X ∝ t'
X ∝ t'^1/2
Symmetric
Asymmetric
Displacement counts
B
Ordering time
N^3
N^2
Sym. σ = 8
Sym. σ = 16
Sym. σ = 32
Sym. σ = 64
Asym.σ = 8
Asym.σ = 16
Asym.σ = 32
Asym.σ = 64
Scaled number of units N/σ
C
Asymmetric
Symmetric
Population
Ordering time
D
Distortion χ
T=500
T=2500
T=8000
Time

moving to the right or to the left. In contrast, for the case of an asymmetric neighborhood function, the mean distance obeys

$$\langle X(\hat{t}) \rangle \propto \hat{t}, \tag{5.2}$$

which indicates that the topological defect behaves like a drift motion along the positive direction of the asymmetric function. This drift motion arises from the asymmetry of the neighborhood function. Thus, equations 5.1 and 5.2 suggest that to eliminate the defect from the N-unit array, the required actual step $\hat{t}$ in which the defect moves is proportional to $O(N^2)$ and $O(N)$ for symmetric and asymmetric neighborhood functions, respectively. For the topological defect to move, some unit near this defect needs to be a winner so that the reference vector associated with the defect is updated. In general, the probability that the unit near the defect becomes a winner depends on both the form of the distribution of reference vectors of all units and the probability density of the input vectors. As a first approximation, however, it is reasonable to estimate that this probability of becoming a winner is proportional to $\frac{1}{N}$ when the input vectors are randomly selected from a uniform distribution. Therefore, the probability that the defect motion occurs is of the order $O(1/N)$. In other words, the total steps required to move the defect out of the feature map are proportional to the system size $O(N)$. Considering the above points, the convergence time can be estimated by $O(N^3)$ and $O(N^2)$ with a constant neighborhood range.

Several conventional ways have been used to avoid the slow convergence of the learning process in the presence of a kink state, such as choosing a suitable set of initial reference vectors or rearranging the sequence of input vectors. Further, it is well known that using a neighborhood function with a large width is effective in creating an ordered map from a very random initial condition. This is partially because a kink state is likely to appear with a narrow neighborhood function. However, the large width tends to average out the fine, detailed, ordered structures. Therefore, in many cases, the width of the neighborhood function is initially set to be large, such as half the width of the array of units, and is gradually decreased to a small final value. In Figure 6B, we investigate the dependency of ordering time on the width of the neighborhood function under the existence of a single kink. Figure 6B shows the ordering time as a function of scaled number of units N, that is, $(N/\sigma)^k$, where $k = 2$ for asymmetric neighborhood function and $k = 3$ for symmetric one. Thus, the large σ reduces the effective number of units N, independent of the shape of the neighborhood function. In other words, this means that if the unit size N is fixed, the ordering time is proportional to $(\frac{1}{\sigma})^k$. What is especially important is that even if the ratio, $\frac{N}{\sigma}$, is kept a constant value, the ordering time with an asymmetric neighborhood function is only about one-tenth that with a symmetric one. This result implies that the combined use of the asymmetric neighborhood function and the method

Table 1: Summary of the Scaling Exponent of Ordering Time to the Total Number of Units N for Various Neighborhood Functions.

	Symmetric	Asymmetric
Gaussian	2.992 ± 0.001	1.912 ± 0.004
Tent	2.980 ± 0.001	1.912 ± 0.003
Parabola	2.977 ± 0.0004	1.919 ± 0.003

Notes: Tent function is described by $h(r) = [1 - r/\sigma]_+$ and parabola function is described by $h(r) = [1 - (r/\sigma)^2]_+$, where $[x]_+ = x$ if $x \geq 0$ and $[x]_+ = 0$ elsewhere. The simulation parameters are the same as in Figure 2D.

of adjusting the width of the neighborhood function is more effective for achieving a fast ordering process than using either technique by itself.

Although the gaussian neighborhood function is commonly used in SOM, some studies show that the shape of the neighborhood function is important for fast learning. We have examined the performance of the ordering process with various types of neighborhood functions, and the results are summarized in Table 1. Table 1 shows that the result found with the gaussian function quantitatively holds in cases of other types of neighborhood functions, such as a parabola and tent functions. However, in the case of a nonconvex neighborhood function, the dynamical properties of the ordering process drastically change owing to the presence of many metastable states. Erwin et al. (1992) have proved that in the case of nonconvex neighborhood functions, there are metastable states in which the system gets trapped during the learning process. In fact, in the case of the concave exponential neighborhood function defined in Erwin's et al.'s letter, the system is trapped in a metastable state and cannot reach a completely ordered state. Consequently, to realize the fast ordering process, the neighborhood function should be convex over most of its support.

In order to investigate the dynamical properties of the topological defect, we have considered a simple situation that a single topological defect exists around the center of the feature map as an initial condition. However, when the initial reference vectors are set randomly, the total number of topological defects appearing in the feature map is not generally equal to one. Therefore, we need to consider the statistical distribution of the ordering time, because the total number of the topological defects and the convergence process depend generally on the initial conditions. Figure 6C shows the distribution of the ordering time in 1D SOM, when the initial reference vectors are randomly selected from the uniform distribution [0, 1]. The arrows indicate the average ordering time in the case of the single-kink initial condition. For the asymmetric neighborhood function, the distribution of the ordering time has a single sharp peak near the value of the average ordering time in the case of a single kink initial condition. In contrast, for the symmetric neighborhood function, the distribution of the ordering time has double peaks and large variations. By examining the time development of the topological order η, it is found that the first higher, sharper peak results from the fast ordering process in the absence of the kink state. The second lower, broader peak arises from the slow process in which some kinks appear during learning. In fact, the average position of this broad peak ($\geq$7500 steps) is around 1.59×10^4 steps, which is close to the ordering time in the case of the single kink initial condition (1.70×10^4 steps). So although the fast ordering process is observed in some successful cases (lucky initial conditions), the averaged behavior of the ordering process can be qualitatively well described by the results obtained in the case of a single kink initial state.

Finally, we briefly consider the optimal value of the flip period T in the improved asymmetric neighborhood function. If T is too short, the system cannot eliminate the topological defects, because the total number of the learning steps within the same flip period is too short to move the defects out of the edge of the array of units. Therefore, the value T is required to be longer than some critical value for ordering the feature map. On the other hand, if T is too long, few flips of the direction of asymmetric function occur, which tends to interfere with the formation of the complete feature map without distortion. Taking account of the above two extreme cases, it is expected that an optimal value of the flip period T exists. In fact, this is confirmed in Figure 6D, which shows the time development of the distortion χ for several values of the flip period T. As a result of many simulations, we find that the optimal value of flip period T is about 2500, which is almost equivalent to the typical ordering time of the feature map.

Acknowledgments

We thank Koji Kurata and Shigeru Shinomoto for helpful discussions and comments on the manuscript. We are also graceful to Kei Takahashi for preparing the preliminary simulation data. This work was supported by CREST (Core Research for Evolutional Science and Technology) of Japan Science and Technology Corporation (JST) and by Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan: Grants 18047014, 18019019, and 18300079.

References

Der, R., Herrmann, M., & Villmann, T. (1997). Time behavior of topological ordering in self-organizing feature mapping. *Biol. Cybern., 77*(6), 419–427.

Erwin, E., Obermayer, K., & Schulten, K. (1992). Self-organizing maps: Stationary states, metastability and convergence rate. *Biol. Cybern., 67*(1), 35–45.

Geszti, T., Csabai, I., Cazokó, F., Szakács, T., Serneels, R., & Vattay, G. (1990). Dynamics of the Kohonen map. In *Statistical Mechanics of Neural Networks: Proceedings of the XIth Sitges Conference* (pp. 341–349). New York: Springer-Verlag.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in cat's visual cortex. *J. Physiol., London, 160*(1), 106–154.

Hubel, D. H., & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in monkey striate cortex. *J. Comp. Neurol., 158*(3), 267–294.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybern., 43*(1), 59–69.

Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). New York: Springer-Verlag.

Ritter, H. (1991). Asymptotic level density for a class of vector quantization processes. *IEEE Trans. Neural Netw., 2*(1), 173–175.

Ritter, H., & Schulten, K. (1986). On the stationary state of Kohonen self-organizing sensory mapping. *Biol. Cybern., 54*(2), 99–106.

Takeuchi, A., & Amari, S. (1979). Formation of topographic maps and columnar microstructures in nerve fields. *Biol. Cybern., 35*(2), 63–72.

Villmann, T., & Claussen, J. C. (2006). Magnification control in self-organizing maps and neural gas. *Neural Comp., 18*, 446–469.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in striate cortex. *Kybernetik, 14*(2), 85–100.

Received January 16, 2006; accepted October 16, 2006.