2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Complex-valued Associative Memory with Strong Thresholds

Michimasa Kitahara† and Masaki Kobayashi†

†Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi
4-3-11, Takeda, Kofu, Yamanashi, 400-8511 Japan

Abstract—A Complex-valued Associative Memory (CAM) stores not only training patterns but also their rotated patterns. They reduce noise robustness. In this paper, we introduce strong thresholds to the neurons. Thresholds relieve storing rotated patterns and improve noise robustness. If thresholds are simply used, their effect is absorbed in quantification of complex-valued neurons. So we use strong thresholds to improve noise robustness.

1. Introduction

Complex-valued Associative Memory (CAM) has been studied as an advanced model of neural networks (Jankowski et.al.[1]). The CAM can treat multi-state information and is often used for storing gray scale patterns (Aoki[3]). It is known that the CAM stores not only training patterns but also their rotated patterns (Kuroe[2]). The rotated patterns are typical spurious patterns and reduce the noise robustness of the CAM. In case of K quantification, there exist $K-1$ rotated patterns for each training pattern. In this paper, we introduce thresholds to the complex-valued neurons in order to improve noise robustness. Thresholds interferes storing rotated patterns. However if thresholds are simply used, their effect is absorbed in quantification of complex-valued neurons. We adopt generalized inverse matrix (Aoki[3]) for learning and introduce strong thresholds so that the CAM can overcome the quantification of complex-valued neurons.

We introduce a parameter t to control the strength of thresholds. We call the parameter t the control parameter. By computer simulations, we found out that the control parameter t became more effective as t grew until a certain level and the effect of t was almost constant when t was beyond it. Moreover our experiences showed that strong thresholds improved the noise robustness enough in case of $P/N \leq 0.1$, where P and N are the number of training patterns and neurons. In particular, they are useful in case that P is small and K is large. The condition $P/N < 0.1$ is often enough large for applications. In fact, many applications of associative memories, for example Deguchi and Ishii[4, 5], Hattori and Hagiwara[6], Osana et. al.[7], Yano and Osana[8], Nakada and Osana[9] and Kitahara et. al.[10], require $P/N < 0.1$. Therefore the strong thresholds are useful.

2. Complex-valued Associative Memory

In this section, we briefly describe the CAM.

First, we define complex-valued neurons. Complex-valued neurons receive and produce complex numbers. Let $\mathbf{S}$ be the set $\{s_k = \exp(2k\theta_K\sqrt{-1}); k = 0, 1, \cdots, K-1\}$, where $\theta_K = \pi/K$. A complex-valued neuron takes $s_k \in \mathbf{K}$ closest to the received complex number. We define this map from complex numbers to $\mathbf{S}$ as $f(\cdot)$.

Next, we construct the CAM. Let complex numbers w_{ji} and γ_j be the connection weight from the neuron i to the neuron j and the threshold of neuron j, respectively. The connection weights require the condition $w_{ji} = \overline{w}_{ij}$, where $\overline{w}$ stands for the complex conjugate of w. Let x_i be the state of the neuron i. Then the input sum I_j to the neuron j is defined as follows:

$$I_j = \sum_{i \neq j} w_{ji} x_i + \gamma_j. \quad (1)$$

Finally we describe learning. Let N be the number of neurons. If a vector $\mathbf{x} = (x_1, x_2, \cdots, x_N)^T$ belonging to $\mathbf{S}^N$, where superscript T stands for the transpose, satisfies $x_j = f(I_j)$ for all j, it is called stable. Learning is to find the parameters to make given training pattern vectors stable. If a noisy pattern $\mathbf{x}'$ of training pattern $\mathbf{x}$ is given, it is expected to restore the original pattern $\mathbf{x}$. The ability to recover from noisy patterns is called noise robustness. Noise robustness is one of the most important problems of associative memories.

3. Generalized Inverse Matrix Learning

We describe the generalized inverse matrix learning in case that all thresholds are zero ($\gamma_j = 0$ for all j). Let $\mathbf{x}_p = (x_{p1}, x_{p2}, \cdots, x_{pN})^T$ be the pth training pattern vector. And we consider $N \times P$ training matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_P)$, where P is the number of training patterns. We define $N \times N$ matrix $\mathbf{W}$ as follows:

$$\mathbf{W} = \mathbf{X}(\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*, \quad (2)$$

where the superscript $*$ stands for the complex conjugate transpose. The matrix $(\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*$ is called the generalized inverse matrix of $\mathbf{X}$. Then it is clear that the matrix $\mathbf{W}$ satisfies the relation $\mathbf{W}^* = \mathbf{W}$. Therefore the diagonal elements w_{jj} are real numbers and the (i,j) element w_{ij} of $\mathbf{W}$ satisfies $w_{ij} = \overline{w}_{ji}$. We regard w_{ij} as the connection weights for $i \neq j$. The diagonal elements w_{jj} are never used for updating the neurons. We set $\gamma_j = 0$ for all j. Definitely $\mathbf{WX} = \mathbf{X}$ is true. So we obtain $x_{pj} = \sum_i w_{ji} x_{pi}$ for all p and j. When a training pattern $\mathbf{x}_p$ is given to the CAM, we obtain the following equation for all j:

$$I_j = \sum_{i \neq j} w_{ji} x_{pi} = (1 - w_{jj}) x_{pj}. \tag{3}$$

Therefore all training patterns $\mathbf{x}_p$ are stable unless $w_{jj} \geq 1$ for some j.

For any $k \neq 0$, consider the pattern $s_k \mathbf{x}_p$, which is called the rotated pattern of $\mathbf{x}_p$. When the rotated pattern $s_k \mathbf{x}_p$ is given to the CAM, the input sum I_j to the neuron j is as follows:

$$I_j = \sum_{i \neq j} w_{ji} (s_k x_{pi}) = s_k (\sum_{i \neq j} w_{ji} x_{pi}). \tag{4}$$

Therefore the rotated patterns $s_k \mathbf{x}_p$ are also stable since the training pattern $\mathbf{x}_p$ is stable. This fact is called rotation invariance (Zemel et. al.[11]). As K becomes large, two problems arise. One is that the number $(K-1)P$ of the rotated patterns grows. It causes that there exist more spurious patterns. The other is that there exist rotated patterns in the neighborhood of the training patterns. As K grows, the states s_1 and s_{K-1} approach 1. Therefore the rotated patterns $s_1 \mathbf{x}_p$ and $s_{K-1} \mathbf{x}_p$ approach $\mathbf{x}_p$. These facts reduce the noise robustness of the CAM.

4. Generalized Inverse Matrix Learning with Thresholds

In this section, we introduce the generalized inverse matrix learning for the CAM with thresholds. The thresholds are introduced to make the rotated patterns unstable. It is necessary that the amplitudes of the thresholds are enough large. To control the amplitudes of thresholds, we introduce a parameter t. We call it the control parameter. The control parameter t is a real number. All training pattern vectors $\mathbf{x}_p$ are extended to $\widetilde{\mathbf{x}}_p = (\mathbf{x}_p^T, t)^T$. The learning matrix $\mathbf{X}$ is also extended to the $(N+1) \times P$ matrix $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \cdots, \widetilde{\mathbf{x}}_P)$. Moreover the matrix $\mathbf{W}$ is replaced by the $(N+1) \times (N+1)$ matrix $\widetilde{\mathbf{W}}$ as follows:

$$\widetilde{\mathbf{W}} = \widetilde{\mathbf{X}} (\widetilde{\mathbf{X}}^* \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^*. \tag{5}$$

Then the matrix $\widetilde{\mathbf{W}}$ satisfies $\widetilde{\mathbf{W}} = \widetilde{\mathbf{W}}^*$ and $\widetilde{\mathbf{W}}\widetilde{\mathbf{X}} = \widetilde{\mathbf{X}}$. Take the (j,p) element of $\widetilde{\mathbf{W}}\widetilde{\mathbf{X}} = \widetilde{\mathbf{X}}$, where $j \leq N$, and then we obtain

$$\sum_i w_{ji} x_{pi} + t w_{j(N+1)} = x_{pj}. \tag{6}$$

For all i and j such that $i, j \leq N$ and $i \neq j$, we regard w_{ij} as connection weights. From $\widetilde{\mathbf{W}} = \widetilde{\mathbf{W}}^*$, we find $w_{ij} = \overline{w}_{ji}$. Moreover we regard $t w_{j(N+1)}$ as γ_j. Thus we obtain the connection weights w_{ji} and the thresholds γ_j. Suppose that a training pattern $\mathbf{x}_p$ is put on the CAM. Then the input sum I_j to the neuron j is as follows:

$$I_j = \sum_{\substack{i \leq N \\ i \neq j}} w_{ji} x_{pi} + \gamma_j \tag{7}$$

$$= (1 - w_{jj}) x_{pj}. \tag{8}$$

From $\widetilde{\mathbf{W}} = \widetilde{\mathbf{W}}^*$, we find that w_{jj} is a real number. All training patterns $\mathbf{x}_p$ are stable unless $w_{jj} \geq 1$ for some j. Suppose that a pattern $\mathbf{x}_p$ is stable and a rotated pattern $s_k \mathbf{x}_p$ is put on the CAM. Then the input sum I_j to the neuron j is as follows:

$$I_j = s_k \sum_{\substack{i \leq N \\ i \neq j}} w_{ji} x_{pi} + \gamma_j \tag{9}$$

$$= s_k (\sum_{i \leq N} w_{ji} x_{pi} + \gamma_j - w_{jj} x_{pj} - \gamma_j + \overline{s_k} \gamma_j) \tag{10}$$

$$= s_k (x_{pj} + (\overline{s_k} - 1)\gamma_j - w_{jj} x_{pj}) \tag{11}$$

$$= s_k x_{pj} (1 - w_{jj} + (\overline{s_k} - 1)\gamma_j \overline{x}_{pj}). \tag{12}$$

If the condition

$$|\arg(1 - w_{jj} + (\overline{s_k} - 1)\gamma_j \overline{x}_{pj})| > \theta_K \tag{13}$$

is true for some j, where $\arg(z)$ stands for the phase of complex number z, the rotated pattern $s_k \mathbf{x}_p$ is not stable. Since the pattern $\mathbf{x}_p$ is stable, $1 - w_{jj}$ is a positive real number. So if the amplitude of γ_j is too small compared with $1 - w_{jj}$, the condition (13) is not true. We have to determine the control parameter t such that amplitudes of some γ_j are sufficient large.

In case of $t = 0$, $\widetilde{\mathbf{X}}^* \widetilde{\mathbf{X}} = \mathbf{X}^* \mathbf{X}$ is true. The (i,j) elements of $\mathbf{W}$ and $\widetilde{\mathbf{W}}$ are same for all $i, j \leq N$. Therefore in case of $t = 0$, the proposed generalized inverse matrix learning is the conventional one. The proposed generalized inverse matrix learning is an extension of the conventional one. And we expect that the control parameter t plays a role in determining the scale of thresholds. In case of $t = 1$, the proposed generalized inverse matrix learning is equivalent to that with a constant input neuron which always produces one.

5. Control Parameter and Thresholds

We expect that the larger the control parameter t becomes the larger the thresholds γ_j become. However it is not trivial. We examined the relation between the control parameter and the thresholds by computer

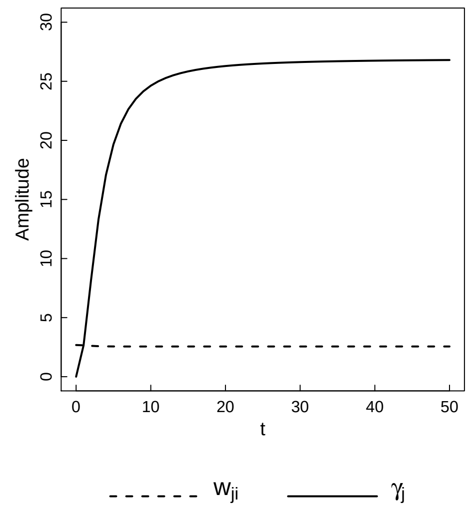

Figure 1: Average amplitudes of w_{ji} and γ_j

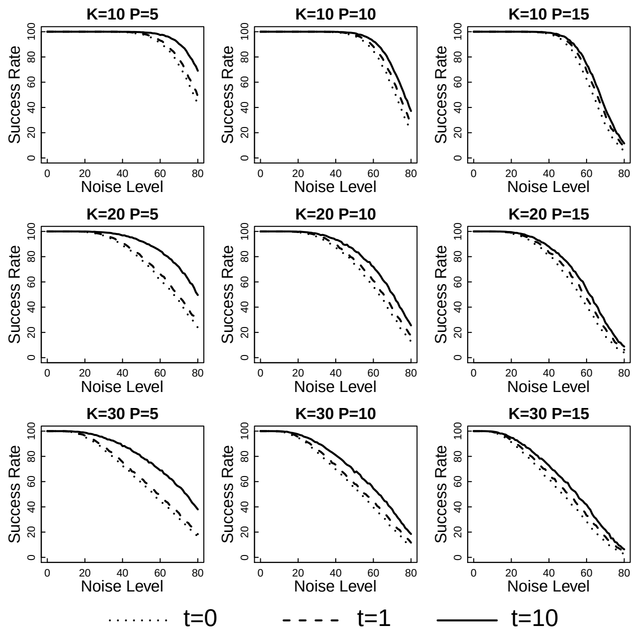

Figure 2: Noise robustness

simulations. The computer simulation was performed under the condition of $N = 100, K = 8$ and $P = 10$. The number of trials for each t was 1000. The training data was generated at random.

Figure 1 shows how the amplitudes of connection weights w_{ji} and thresholds γ_j depend on the control parameter t. The x-axis shows the control parameter t. The y-axis shows the average amplitudes of the connection weights w_{ji} and thresholds γ_j . The average amplitude of the connection weights w_{ji} was almost independent of t. The average amplitude of the thresholds γ_j increased until a certain level of t and was almost constant beyond the level. Now we are interested in the ratio r of the average amplitude of the thresholds γ_j to that of the connection weights w_{ji}. The ratio r is required to be enough large. This experiment implies that the ratio r rapidly increased for small t and was almost constant for enough large control parameter t.

6. Noise Robustness

In this section, we perform computer simulations in order to examine the effect of control parameter t for noise robustness.

The simulation has been carried out under the conditions $N = 100$, $K = 10, 20$ and 30, $P = 10, 20$ and 30, and $t = 0, 1$ and 10. In each condition, 100 training data sets were generated at random. For each training data set and each noise level L, 100 trials were carried out by the following procedure.

1. A training pattern was selected at random and put on the CAM.
2. L neurons were selected and replaced by any states at random.
3. If the CAM recalls the original pattern correctly, we regarded the trial as successful.

The condition $t = 0$ means the case without thresholds. The condition $t = 1$ means the case with a constant input neuron which always produces one. In Fig. 1, the amplitudes of thresholds rapidly increased in case of $t < 10$ and were almost constant in case of $t > 10$. We can expect that the effect of thresholds is independent of t in case of $t \geq 10$. In our experiences, actually, the simulation results in cases of $t = 10$ and 20 were almost same.

Figure 2 shows the simulation results. The x-axis shows the noise level. The y-axis shows the success rate. In the simulation results, we can find the following facts.

1. In all cases, the noise robustness was more improved as t increased.
2. In case of $t = 1$, noise robustness was just a bit improved.
3. The differences in improvement of $t = 1$ and 10 were larger as P was smaller or K was larger.

Our experiences show that the strong thresholds work well especially in case of $P/N \leq 0.1$. The ability is often effective for many applications of associative memories as Table 1. All of them require $P/N < 0.1$.

Applications	P/N
Deguchi and Ishii[4]	0.04
Deguchi and Ishii[5]	0.075
Hattori and Hagiwara[6]	0.04
Osana et. al.[7]	0.011
Yano and Osana[8]	0.009
Nakada and Osana[9]	0.008
Kitahara et. al.[10]	0.008

Table 1: P/N of some applications

7. Conclusion

In this paper, we proposed generalized inverse matrix learning for the CAM with thresholds to improve noise robustness. The CAM without thresholds stores many rotated patterns, which reduces the noise robustness. The thresholds can prevent the CAM from storing rotated patterns. However, if the thresholds are too weaker compared with the connection weights, we cannot expect the sufficient effect of the thresholds. So we introduce the control parameter t to enlarge the thresholds. By the computer simulations, we found the following results for the ratio r of the average amplitude of the thresholds to that of the connection weights.

1. In case of small t, the ratio r increased as t grows.
2. In case of large t, the ratio r was almost constant and independent to t.

Moreover, we carried out the computer simulations to test noise robustness. The simulation results showed that our proposal improved noise robustness. The improvement was better, as the number P of training patterns was smaller or the number K of states of neuron was larger.

References

[1] S. Jankowski, A. Lozowski and J. M. Zurada, "Complex-valued multistate neural associative memory," *IEEE Trans. Neural Networks*, vol.7, pp.1491–1496, 1996.

[2] Y. Kuroe, "A Model of complex-valued associative memories and its dynamics," *Complex-Valued Neural Networks, Theories and Applications*, Ed. A.Hirose, World Scientific, pp.57–79, 2003.

[3] H. Aoki, M. R. Azimi-Sadjadi and Y. Kosugi, "Image association using a complex-valued associative memory model," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol.E83-A, pp.1824–1832, 2000.

[4] T. Deguchi and N. Ishii, "Control of associative dynamics by matching features in chaotic neural network," *Systems and Computers in Japan*, vol.27, no.5, pp.47–54, 1996.

[5] T. Deguchi and N. Ishii, "Search for cyclic patterns by a chaotic neural network," *Systems and Computers in Japan*, vol.29, no.11, pp.1–8, 1998.

[6] M. Hattori and M. Hagiwara, "Multimodule associative memory for many-to-many associations," *Neurocomputing*, vol.19, pp.99–119, 1998.

[7] Y. Osana, M. Hattori and M. Hagiwara, "Chaotic bidirectional associative memory," *Proceedings of IEEE International Conference on Neural Networks*, pp.816–821, 1996.

[8] Y. Yano and Y. Osana, "Chaotic complex-valued bidirectional associative memory," *Proceedings of IEEE and INNS International Joint Conference on Neural Networks*, pp.433–438, 2009.

[9] M. Nakada and Y. Osana, "Chaotic complex-valued associative memory," *2007 International Symposium on Nonlinear Theory and its Applications NOLTA'07*, pp.493–496, 2007.

[10] M. Kitahara, M. Kobayashi and M. Hattori, "Chaotic rotor associative memory," *2009 International Symposium on Nonlinear Theory and its Applications NOLTA'09*, pp.399–402, 2009.

[11] R. S. Zemel, C. K. I. Williams and M. C. Mozer, "Lending direction to neural networks," *Neural Networks*, pp.503–512, 1995.