# Plateau in a Polar Variable Complex-valued Neuron

Tohru Nitta
*National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan*

Keywords: Complex-valued neuron, Coordinate, Learning, Unidentifiable model, Singularity.

Abstract: In this paper, the characteristics of the complex-valued neuron model with parameters represented by polar coordinates (called *polar variable complex-valued neuron*) are investigated. The main results are as reported below. The polar variable complex-valued neuron is unidentifiable: there exists a parameter that does not affect the output value of the neuron and one cannot identify its value. The plateau phenomenon can occur during learning of the polar variable complex-valued neuron: the learning error does not decrease in a period. Furthermore, it is suggested by computer simulations that a single polar variable complex-valued neuron has the following characteristics: (a) Unidentifiable parameters (singular points) degrade the learning speed. (b) A plateau can occur during learning. When the weight is attracted to the singular point, the learning tends to be stuck.



## 1 INTRODUCTION

A complex-valued neural network is a general neural network with parameters such as weight and a threshold value extending from real to complex numbers. Complex-valued neural networks are suitable for information processing of complex-valued data or two-dimensional data (Hirose, 2006; Nitta, 2009; Hirose, 2013).

Conventionally, an approach for real numbers must be applied to a real part and imaginary part separately, whereas a complex-valued neural network allows direct data processing. It is also advantageous because good-natured behavior of the complex number to rotation can be taken automatically. Consequently, some properties that are intrinsic to a complex-valued neural network have been clarified (Nitta, 2008).

Learning models have been studied by relation with singular points lately (Amari et al., 2006; Wei et al., 2008; Cousseau et al., 2008; Nitta, 2013). For example, learning models with hierarchic structures or a symmetric property on exchange of weights, such as hierarchical neural networks and mixture models, have singular points, mostly. It has been proved that singular points affect the learning dynamics of learning models, and that they can cause a standstill in learning.

Properties of the singular points of complex-valued neuron constituting a complex-valued neural network are investigated in this paper. A usual neuron with real-valued weights and a real-valued threshold is designated as a *real-valued neuron*. A neural network comprising real-valued neurons is designated as a *real-valued neural network*. Generally, a complex number can be expressed in two ways: using orthogonal coordinates and with polar coordinates. A complex-valued neuron whose parameters (weight and threshold) are expressed with an orthogonal coordinate is designated as an *orthogonal variable complex-valued neuron*, whereas a complex-valued neuron whose parameters are expressed using a polar coordinate is designated as a *polar variable complex-valued neuron*. The literature (Hirose et al., 2001; Kawata and Hirose, 2003; Hirose et al., 2004; Hirose, 2006) includes numerous explanations of complex-valued neural network models comprising polar variable complex-valued neurons and their applications.

This paper demonstrates that a polar variable complex-valued neuron is unidentifiable. Mathematical indications show that a plateau phenomenon can occur during learning (Nitta, 2010). Then it is suggested experimentally that (a) unidentifiable parameters (singular points) degrade the learning speed. (b) A plateau can occur during learning.

Properties related to the singular points of a polar variable complex-valued neuron are investigated analytically in Section 2. Then, using computer simulations, the kind of effect the singular point has on the learning dynamics of a polar variable complex-valued


Nitta T..
Plateau in a Polar Variable Complex-valued Neuron.
DOI: 10.5220/0004887905260531
In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence* (ICAART-2014), pages 526-531
ISBN: 978-989-758-015-4

neuron is investigated in Section 3. Finally, this paper is concluded. Future topics are described in Section 4.

## 2 ANALYSIS

The singularity of a polar variable complex-valued neuron is investigated analytically in this section.

### 2.1 Unidentifiability of a Polar Variable Complex-valued Neuron

In this section, it is shown that a polar variable complex-valued neuron is unidentifiable. A connected set comprising parameter values for which polar variable complex-valued neurons take an identical output value is designated as a *critical set*, whereas points on a critical set are regarded as singular points in this paper. Only connected sets were employed as analysis objects because an unconnected set is considered to have no bad effect on learning dynamics.

The following polar variable complex-valued neuron of $N$ input is assumed. Then output value $v$ is defined as

$$v = f_C\left(\sum_{k=0}^{N} r_k \exp[i\theta_k]\cdot z_k\right) \in C, \qquad (1)$$

where $C$ stands for the set of complex numbers, $z_k \in C$ signifies the $k$-th input signal, $r_k \exp[i\theta_k] \in C$ denotes weight to the $k$-th input signal ($r_k \in R$ represents the amplitude and $\theta_k \in R$ is phase where $R$ is the set of real numbers) $(1 \le k \le N)$, $i = \sqrt{-1}$, $z_0 \equiv 1$, $r_0 \exp[\theta_0] \in C$ represents the threshold of a complex-valued neuron ($r_0 \in R$ is amplitude and $\theta_0 \in R$ is phase). In addition, $f_C : C \to C$ is an activation function.

In the polar variable complex-valued neuron described above, if the amplitude parameter $r_k$ is equal to zero for some $0 \le k \le N$, then [weight $\times$ input] = $r_k \exp[i\theta_k]\cdot z_k = 0$ holds, and no value of $\theta_k$ affects the output value $v$ of a complex-valued neuron. That is, one cannot identify the value of the phase parameter $\theta_k$ uniquely when the amplitude parameter $r_k$ is equal to zero. Therefore, it is verified that the phase $\theta_k$ is an unidentifiable parameter and a polar variable complex-valued neuron has an unidentifiable nature.

Next, the critical set of the polar variable complex-valued neuron described above is specifically determined. First, let

$$M \stackrel{\text{def}}{=} \{(r,\Theta) \in R^{N+1} \times R^{N+1}\}, \qquad (2)$$

$$r \stackrel{\text{def}}{=} \begin{bmatrix} r_0 \\ \vdots \\ r_N \end{bmatrix} \in R^{N+1}, \qquad (3)$$

$$\Theta \stackrel{\text{def}}{=} \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_N \end{bmatrix} \in R^{N+1}, \qquad (4)$$

where $M$ is a parameter space that specifies the polar variable complex-valued neuron described above. Then, for any $(r',\Theta') \in M$ and any $0 \le k \le N$, let

$$\begin{aligned} C_k(r',\Theta') \stackrel{\text{def}}{=} \{(r,\Theta) \in M \mid\ & r_0 = r'_0, \cdots, \\ & r_{k-1} = r'_{k-1}, r_k = 0, r_{k+1} = r'_{k+1}, \\ & \cdots, r_N = r'_N, \theta_0 = \theta'_0, \cdots, \\ & \theta_{k-1} = \theta'_{k-1}, \theta_{k+1} = \theta'_{k+1}, \cdots, \\ & \theta_N = \theta'_N\}. \end{aligned} \qquad (5)$$

Then the critical set of the polar variable complex-valued neuron described above, $C(r',\Theta')$ is given as

$$C(r',\Theta') = \cup_{k=0}^{N} C_k(r',\Theta'). \qquad (6)$$

Actually, for any $(r,\Theta) \in C(r',\Theta')$, there exists some $k$ such that $(r,\Theta) \in C_k(r',\Theta')$. Therefore considering $r_k = 0$,

$$\begin{aligned} v &= f_C\big(r'_0 \exp[i\theta'_0]z_0 + \cdots + r'_{k-1}\exp[i\theta'_{k-1}]z_{k-1} \\ &\quad + r_k \exp[i\theta_k]z_k + r'_{k+1}\exp[i\theta'_{k+1}]z_{k+1} + \cdots \\ &\quad + r'_N \exp[i\theta'_N]z_N\big) \\ &= f_C\big(r'_0 \exp[i\theta'_0]z_0 + \cdots + r'_{k-1}\exp[i\theta'_{k-1}]z_{k-1} \\ &\quad + 0 + r'_{k+1}\exp[i\theta'_{k+1}]z_{k+1} + \cdots \\ &\quad + r'_N \exp[i\theta'_N]z_N\big) \end{aligned} \qquad (7)$$

holds, and $v$ remains constant irrespective of $\theta_k \in R$.

### 2.2 Learning Dynamics Near the Singular Point of a Polar Variable Complex-valued Neuron

Learning dynamics near the singular point of a polar variable complex-valued neuron is investigated using the analysis procedure of reference (Amari et al., 2006).

A polar variable complex-valued neuron defined in section 2.1 is adopted as an analysis object. The activation function $f_C$ is taken as a linear function for simplicity.

$$f_C(z) = z, \qquad z = x + iy. \qquad (8)$$

Error is defined as $E = (1/2)|t-v|^2$ ($t \in C$ is teacher signal and $v \in C$ is an actual output value).

In fact, $E$ is a complex function, but it takes only a real value as function values, and is not regular as a complex function. That is, $E$ is not complex differentiable. Nevertheless, it is possible to derive a learning rule by considering the partial differential. In that case, the learning dynamics of a complex-valued neuron change according to whether the parameter (weight and threshold) is considered as orthogonal coordinates system, or it is considered as a polar coordinate system. Although error function $E$ was discussed, the same argument applies to the complex differentiability of activation function $f_C$ (See (Hirose, 2006, pp.18-22) for details).

A learning rule is derived as follows using the steepest descent method: For any $0 \le k \le N$,

$$\begin{aligned} \Delta r_k(n) &\stackrel{\text{def}}{=} r_k(n+1) - r_k(n) \\ &= -\varepsilon \cdot \frac{\partial E}{\partial r_k} \\ &= \varepsilon \cdot Re\left[\,\overline{\delta} \cdot z_k \cdot \exp[i\theta_k(n)]\right], \end{aligned} \quad (9)$$

$$\begin{aligned} \Delta\theta_k(n) &\stackrel{\text{def}}{=} \theta_k(n+1) - \theta_k(n) \\ &= -\varepsilon \cdot \frac{\partial E}{\partial \theta_k} \\ &= -\varepsilon \cdot r_k(n) \cdot Im\left[\,\overline{\delta} \cdot z_k \cdot \exp[i\theta_k(n)]\right], \end{aligned} \quad (10)$$

where $\delta \stackrel{\text{def}}{=} t - v$, $\bar{z}$ is a complex conjugate of complex $z$, and $n$ is a variable that represents the number of learning cycles. For example, $r_k(n)$ expresses the value of parameter $r_k$ after finishing learning of $n$ times.

A learning rule of a single complex-valued neuron whose weight is expressed on a polar coordinate is derived using the steepest descent method in the reference (Hirose, 2006, pp.59-64). Because the following nonlinear function (amplitude - phase type activation function) is used as the activation function of the complex-valued neuron concerned, difference in expression has occurred from the learning rule derived in this paper:

$$f_{ap}(u) = \tanh(|u|) \cdot \exp[i \cdot \arg(u)],\ u \in C. \quad (11)$$

For any $0 \le k \le N$, define

$$M_{r_k} \stackrel{\text{def}}{=} \{\ (r,\Theta) \in M \ \mid\ \Delta r_k = 0\ \}, \quad (12)$$

$$M_{\theta_k} \stackrel{\text{def}}{=} \{\ (r,\Theta) \in M \ \mid\ \Delta\theta_k = 0\ \}. \quad (13)$$

Then learning rules (Eqs. (9) and (10)) yield

$$M_{r_k} = \left\{(r,\Theta) \in M \,\middle|\, Re\left[\,\overline{\delta} z_k \cdot \exp[i\theta_k]\right] = 0\right\}, \quad (14)$$

$$M_{\theta_k} = \left\{(r,\Theta) \in M \,\middle|\, r_k \cdot Im\left[\,\overline{\delta} z_k \cdot \exp[i\theta_k]\right] = 0\right\}. \quad (15)$$

Table 1: Training patterns used in the experiment.

| | Input | Output |
|---|---|---|
| Pattern 1 | $1.0$ | $0.5i$ |
| Pattern 2 | $0.5 - 0.5i$ | $-0.5 + 0.5i$ |
| Pattern 3 | $-0.5 - 0.5i$ | $1.0 - 0.5i$ |

Next, the behavior of learning near singular points is investigated. Near singular point $r_k = 0$ ( $k = 0, \cdots, N$), for $k = 0, \cdots, N$, Eqs. (9), (10) yield,

$$\Delta r_k = \varepsilon \cdot Re\left[\,\overline{\delta} z_k \cdot \exp[i\theta_k]\right], \quad (16)$$

$$\Delta\theta_k \fallingdotseq 0. \quad (17)$$

Therefore, the velocity of change of amplitude $r_k$ ( $k = 0, \cdots, N$) is higher than the velocity of phase $\theta_k$ ( $k = 0, \cdots, N$), and a state is attracted to the sub-manifold $\cap_{k=0}^{N} M_{r_k}$ (State $\Delta r_k \fallingdotseq 0$ $(k = 0, \cdots, N)$ is approached). That is, an equilibrium state $\cap_{k=0}^{N}\{M_{r_k} \cap M_{\theta_k}\}$ is reached, and consequently parameter $(r,\Theta) \in M$ will change only slightly. This is a plateau phenomenon in a learning curve, which is the same as that in learning dynamics near singular points of a real-valued neural network, as demonstrated in an earlier study (Amari et al., 2006).

# 3 EXPERIMENT

In this section, behavior of learning near the singular points of a polar variable complex-valued neuron is investigated experimentally.

A polar variable complex-valued neuron of one input is used for simplicity. The activation function $f_C$ is assumed as a linear function:

$$f_C(z) = z, \qquad z = x + iy. \quad (18)$$

We assume that the threshold $w_0 = r_0 \cdot \exp[i\theta_0] \equiv 0$. That is, the learnable parameter is only one weight $w_1 = r_1 \cdot \exp[i\theta_1]$. The general steepest descent method (Eqs. (9), (10)) was used in learning. The learning rate was set to 0.5. Training patterns are of three types, as shown in Table 1. Learning was judged to converge, and terminated when the learning error $(1/2)|t - v|^2$ dropped to 0.0001 or less ($t$ is a teacher signal and $v$ is the actual output value of a polar variable complex-valued neuron).

At the singular point of the polar variable complex-valued neuron described above, $r_1 = 0$ (the amplitude of the weight $w_1$ is zero). Therefore, the initial value of $r_1$ was set to 0.00001, assuming the case in which learning was started near singular point $r_1 = 0$ (Case 1 of Table 2). Moreover, initial value $r_1 = 1.0$ was adopted assuming that learning started

Table 2: Initial values of amplitude of weight. Case 1: Learning is started from near the singular point. Case 2: Learning is started from off the singular point.

| | $r_1$ |
|---|---|
| Case 1 | 0.00001 |
| Case 2 | 1.0 |

Table 3: Initial values of phase $\theta_1$ of weight $w_1$.

| Case | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Initial value | 0 | $\frac{\pi}{4}$ | $\frac{\pi}{2}$ | $\frac{3\pi}{4}$ | $\pi$ | $\frac{5\pi}{4}$ | $\frac{3\pi}{2}$ | $\frac{7\pi}{4}$ |

from a point distant from the singular point $r_1 = 0$ (Case 2 of Table 2). The initial value of the phase $\theta_1$ of the weight $w_1$ was chosen from the eight types presented in Table 3.

The experimental results are presented in Table 4. The average numbers of training cycles starting from near the singular point were 1.52 times ($\fallingdotseq$83.88/55.13), 1.71 times ($\fallingdotseq$55.75/32.63) those of starting off from the singular point for training patterns 1 and 2, respectively. The average number of training cycles starting from near the singular point in training pattern 3 was 1.05 times ($\fallingdotseq$34.50/33.00) that starting off from the singular point. Thus, we could realize from the above results that the average learning speed of the polar variable complex-valued neuron starting from near the singular point is about 1.5 - 1.7 times slower than or comparable to that starting off from the singular point.

When starting from near the singular point, we observed a plateau phenomenon in the case 5 for the training pattern 1 (Fig. 1). A standstill in learning occurred from 1st to 110th cycle. The transitions of the amplitude $r_1$ and the phase $\theta_1$ of the weight $w_1$ are shown in Figs. 2 and 3, respectively. As shown in Fig. 4, the speed of change of the amplitude is faster than that of the phase: $\Delta r_1 > \Delta\theta_1$. And also, the phase $\theta_1$ changed little up to around 90th learning cycle: $\Delta\theta_1 \fallingdotseq 0$. These observation results have agreed with the theoretical results presented in Section 2.2. The amplitude of the weight was attracted to singular point 0 from 1st to around 100th cycle.

It seems at a glance from Fig. 1 that the error remains completely unchanged and a plateau occurs from 1st to 110th learning cycle. However, the actual data says that this is not true. For a fact, the error remains completely unchanged during 1-40, 42-46, 48-49, and 54-55 learning cycles, respectively: plateau occurs in each period. In other periods, the error decreases albeit only slightly. However, roughly speaking, we could say that a quasi-plateau occurs from 1st to 110th cycle.

Experimental results suggest the following for singular points and learning dynamics of polar variable complex-valued neurons. (a) When learning is started near the singular point, a mostly greater than average number of training cycles is required compared with the case in which learning is started from off the singular point. (b) A plateau can occur during learning. When the weight is attracted to the singular point, the learning speed tends to be stuck.

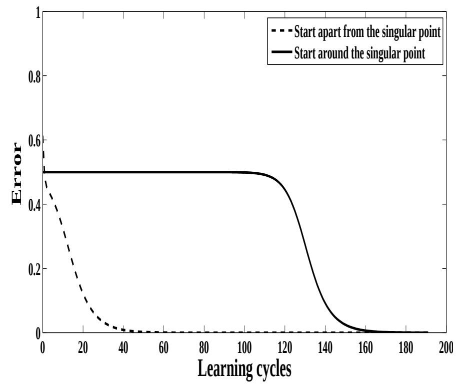


Figure 1: A Learning curve (Training Pattern 1, Case 5).

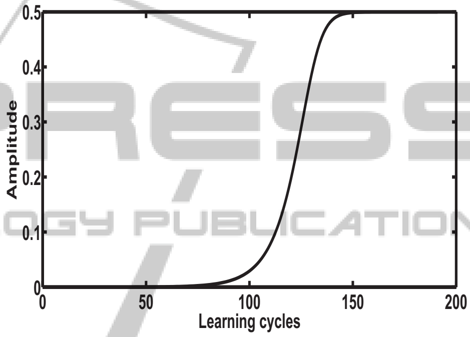


Figure 2: Transition of the amplitude $r_1$ of the weight $w_1$ (Training Pattern 1, Case 5, starting from near the singular point).

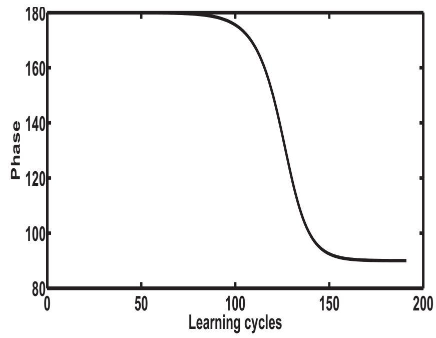


Figure 3: Transition of the phase $\theta_1$ of the weight $w_1$ (Training Pattern 1, Case 5, starting from near the singular point).

Table 4: Number of training cycles necessary for convergence. Case number implies those presented in Table 3 (the initial value of the phase $\theta_1$ of the weight $w_1$).

(a) Training pattern 1

| Case | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Average |
|---|---|---|---|---|---|---|---|---|---|
| Start around the singular point ($r_1 = 0.00001$) | 191 | 66 | 12 | 66 | 192 | 66 | 12 | 66 | 83.88 |
| Start apart from the singular point ($r_1 = 1.0$) | 74 | 61 | 12 | 61 | 74 | 73 | 13 | 73 | 55.13 |

(b) Training pattern 2

| Case | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Average |
|---|---|---|---|---|---|---|---|---|---|
| Start around the singular point ($r_1 = 0.00001$) | 30 | 37 | 119 | 37 | 30 | 37 | 119 | 37 | 55.75 |
| Start apart from the singular point ($r_1 = 1.0$) | 33 | 45 | 38 | 31 | 0 | 31 | 38 | 45 | 32.63 |

(c) Training pattern 3

| Case | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Average |
|---|---|---|---|---|---|---|---|---|---|
| Start around the singular point ($r_1 = 0.00001$) | 37 | 36 | 32 | 33 | 37 | 36 | 32 | 33 | 34.50 |
| Start apart from the singular point ($r_1 = 1.0$) | 36 | 31 | 29 | 29 | 32 | 38 | 34 | 35 | 33.00 |



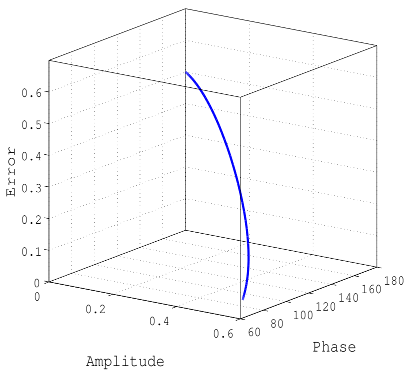


Figure 4: Transition of the error with respect to the amplitude $r_1$ and the phase $\theta_1$ of the weight $w_1$ (Training Pattern 1, Case 5, starting from near the singular point).

## 4 CONCLUSIONS

The singularity of a polar variable complex-valued neuron is investigated in this paper. The following results are obtained. (a) A polar variable complex-valued neuron is unidentifiable. That is, there exists a parameter that does not affect the output value of the neuron, and as a result one cannot identify its value. (b) A plateau can occur during learning of a polar variable complex-valued neuron. In the plateau period, the learning error does not decrease. (c) Singular points (or critical points) degrade the learning speed. When using polar variable complex-valued neurons, one should pay attention to these properties.

It is reported that the expectation of generalization error in cases where true parameters are unidentifiable is greater than in cases where true parameters are identifiable in a three-layer real-valued neural network (Fukumizu, 1999). Properties peculiar to singular points, including whether the generalization error of a polar variable complex-valued neuron deteriorates, are interesting subjects for future study.

## ACKNOWLEDGEMENTS

The author would like to give special thanks to the anonymous reviewers for valuable comments.

## REFERENCES

Amari, S., Park, H., and Ozeki, T. (2006). Singularities affect dynamics of learning in neuromanifolds. *Neural Computation*, 18(5):1007–1065.

Cousseau, F., Ozeki, T., and Amari, S. (2008). Dynamics of learning in multilayer perceptrons near singularities. *IEEE Trans. Neural Networks*, 19(8):1313–1328.

Fukumizu, K. (1999). Generalization error of the maximum likelihood estimation of layered models. In *Proc. 1999*

*Workshop on Information-Based Induction Sciences (IBIS'99)*, pages 221–226. (In Japanese).

Hirose, A. (2006). *Complex-valued neural networks, Series on Studies in Computational Intelligence*, volume 32. Springer-Verlag.

Hirose, A. (2013). *Complex-valued neural networks: advances and applications*. IEEE Press-Wiley.

Hirose, A., Asano, Y., and Hamano, T. (2004). Mode-utilizing developmental learning based on coherent neural networks. In *Proc. Int. Conf. on Neural Information Processing (ICONIP2004)*, pages 116–121, Berlin. Springer.

Hirose, A., Tabata, C., and Ishimaru, D. (2001). Coherent neural network architecture realizing a self-organizing activeness mechanism. In Baba, N., Jain, J. C., and Howlett, R. J., editors, *Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies*, pages 576–580. IOS Press, Tokyo.

Kawata, S. and Hirose, A. (2003). A coherent optical neural network that learns desirable phase values in frequency domain by using multiple optical-path differences. *Opt. Lett.*, 28(24):2524–2526.

Nitta, T. (2008). Complex-valued neural network and complex-valued back-propagation learning algorithm. In Hawkes, P. W., editor, *Advances in Imaging and Electron Physics*, volume 152, pages 153–221. Elsevier, Amsterdam, The Netherlands.

Nitta, T. (2009). *Complex-valued neural networks: utilizing high-dimensional parameters*. Information Science Reference, Pennsylvania.

Nitta, T. (2010). On the singularity of a single complex-valued neuron. *IEICE Trans. Inf. & Syst. D*, J93-D(8):1614–1621. (in Japanese).

Nitta, T. (2013). Local minima in hierarchical structures of complex-valued neural networks. *Neural Networks*, 43:1–7.

Wei, H., Zhang, J., Cousseau, F., Ozeki, T., and Amari, S. (2008). Dynamics of learning near singularities in layered networks. *Neural Computation*, 20:813–843.