Published in *Neural Computation*, Vol.16, No.1, pp.73-97, 2004 as a letter.
(Submitted July 1, 2002; accepted April 11, 2003)

(a) **The title of the article**
Orthogonality of Decision Boundaries in Complex-valued Neural Networks

(b) **The authors' full names**
Tohru Nitta

(c) **Affiliations**
Neuroscience Research Institute,
National Institute of Advanced Industrial Science and Technology (AIST),
AIST Tsukuba Central 2, 1-1-1 Umezono, Tsukuba-shi, Ibaraki, 305-8568 Japan.
E-mail: tohru-nitta@aist.go.jp

Orthogonality of Decision Boundaries in Complex-valued Neural Networks

Tohru Nitta
National Institute of Advanced Industrial Science and Technology (AIST),
Tsukuba Central 2, 1-1-1 Umezono, Tsukuba-shi, Ibaraki, 305-8568 Japan.

This paper presents some results of an analysis on the decision boundaries of complex-valued neural networks whose weights, threshold values, input and output signals are all complex numbers. The main results may be summarized as follows. (a) A decision boundary of a single complex-valued neuron consists of two hypersurfaces which intersect orthogonally, and divides a decision region into four equal sections. The XOR problem and the detection of symmetry problem which cannot be solved with 2-layered real-valued neural networks, can be solved by 2-layered complex-valued neural networks with the orthogonal decision boundaries, which reveals a potent computational power of complex-valued neural nets. Furthermore, the fading equalization problem can be successfully solved by the 2-layered complex-valued neural network with the highest generalization ability. (b) A decision boundary of a three-layered complex-valued neural network has the orthogonal property as a basic structure, and its two hypersurfaces approach orthogonality as all the net inputs to each hidden neuron grow. In particular, most of the decision boundaries in the 3-layered complex-valued neural network intersect orthogonally when the network is trained using the Complex-BP algorithm. As a result, the orthogonality of the decision boundaries improves its generalization ability. (c) Furthermore, the average of the learning speed of the Complex-BP is several times faster than that of the Real-BP. The standard deviation of the learning speed of the Complex-BP is smaller than that of the Real-BP. It seems that the complex-valued neural network and the related algorithm are natural for learning of complex-valued patterns for the above reasons.

1 Introduction

It is expected that complex-valued neural networks, whose parameters (weights and threshold values) are all complex numbers, will have applications in fields dealing with complex numbers such as telecommunications, speech recognition and image processing with the Fourier transformation. When using the existing method for real numbers, we must apply the method individually to their real and imaginary parts. On the other hand, complex-valued neural networks allow us to directly process data. Moreover complex-valued neural networks enable us to automatically capture good rotational behavior of complex numbers. Actually, we can find some applications of the complex-valued neural

networks to various fields such as telecommunications and image processing in the literature (Miyauchi, Seki, Watanabe, & Miyauchi, 1992, 1993; Watanabe, Yazawa, Miyauchi, & Miyauchi, 1994; KES, 2001, 2002; ICONIP, 2002). The fading equalization technology is an application domain suitable for the complex-valued neural network, which will be described in Section 4.3.

The back-propagation algorithm (called here, *Real-BP*) (Rumelhart, Hinton, & Williams, 1986a, 1986b) is an adaptive procedure which is widely used in training a multi-layer perceptron for a number of classification applications in areas such as speech and image recognition. The *Complex-BP* algorithm is a complex-valued version of the Real-BP, which was proposed by several researchers independently in the early 1990's (Kim & Guest, 1990; Nitta & Furuya, 1991; Benvenuto & Piazza, 1992; Georgiou & Koutsougeras, 1992; Nitta, 1993, 1997). The Complex-BP algorithm can be applied to multi-layered neural networks whose weights, threshold values, input and output signals are all complex numbers. This algorithm enables the network to learn complex-valued patterns naturally, and has an *ability to transform geometric figures* as its inherent property, which may be related to the Identity Theorem in complex analysis (Nitta & Furuya, 1991; Nitta, 1993, 1997).

It is important to clarify the characteristics of the complex-valued neural networks in order to promote the real applications. This paper makes clear the differences between the real-valued neural network and the complex-valued neural network by analyzing their fundamental properties from the view of network architectures, and clarifies the utility for the complex-valued neural network which the properties discovered in this paper bring about. The main results may be summarized as follows. (a) A decision boundary of a single complex-valued neuron consists of two hypersurfaces which intersect orthogonally, and divides a decision region into four equal sections. The XOR problem and the detection of symmetry problem which cannot be solved with 2-layered real-valued neural networks, can be solved by 2-layered complex-valued neural networks with the orthogonal decision boundaries, which reveals a potent computational power of complex-valued neural nets. Furthermore, the fading equalization problem can be successfully solved by the 2-layered complex-valued neural network with the highest generalization ability. (b) A decision boundary of a three-layered complex-valued neural network has the orthogonal property as a basic structure, and its two hypersurfaces approach orthogonality as all the net inputs to each hidden neuron grow. In particular, most of the decision boundaries in the 3-layered complex-valued neural network intersect orthogonally when the network is trained using the Complex-BP algorithm. As a result, the orthogonality of the decision boundaries improves its generalization ability. (c) Furthermore, the average of the learning speed of the Complex-BP is several times faster than that of the Real-BP. The standard deviation of the learning speed of the Complex-BP is smaller than that of the Real-BP.

This paper is organized as follows: Section 2 describes the complex-valued neural network and the related Complex-BP algorithm. Section 3 deals with the theoretical analyses of decision boundaries of the complex-valued neural network model. The utility for the complex-valued neural network which the properties discovered in Section 3 bring about are given in Section 4. Section 5 is devoted to the discussion on the results obtained in this paper. Finally, we give some conclusions.

2 The Complex-valued Neural Network

This section describes the complex-valued neural network used in the analysis.

2.1 The Model

First, we will consider the following complex-valued neuron. The input signals, weights, thresholds and output signals are all complex numbers. The net input U_n to a complex-valued neuron n is defined as:

$$U_n = \sum_m W_{nm} X_m + V_n, \tag{1}$$

where W_{nm} is the (complex-valued) weight connecting complex-valued neurons n and m, X_m is the (complex-valued) input signal from complex-valued neuron m, and V_n is the (complex-valued) threshold value of neuron n. To obtain the (complex-valued) output signal, convert the net input U_n into its real and imaginary parts as follows: $U_n = x + iy = z$, where i denotes $\sqrt{-1}$. The (complex-valued) output signal is defined to be

$$f_C(z) = f_R(x) + i f_R(y), \tag{2}$$

where $f_R(u) = 1/(1 + \exp(-u)), u \in \boldsymbol{R}$ ($\boldsymbol{R}$ denotes the set of real numbers), that is, the real and imaginary parts of an output of a neuron mean the sigmoid functions of the real part x and imaginary part y of the net input z to the neuron, respectively. Note that the activation function f_C is not a regular complex-valued function because the Cauchy-Riemann equation does not hold.

A complex-valued neural network consists of such complex-valued neurons described above. The Complex-BP learning rule (Nitta & Furuya, 1991; Nitta, 1993, 1997) has been obtained by using a steepest descent method for such (multi-layered) complex-valued neural networks.

2.2 The Activation Function

In this section, the validity of the activation function defined in expression (2) is described.

In the case of real-valued neural networks, an activation function of real-valued neurons is usually chosen to be a smooth (continuously differentiable) and bounded function such as a sigmoidal function. As some researchers have pointed out (Georgiou & Koutsougeras, 1992; Nitta, 1997; Kim & Adali, 2000; Kuroe, Hashimoto, & Mori, 2002), in the complex region, however, we should recall the Liouville's theorem, which says that if a function G is regular at all $z \in \boldsymbol{C}$ and bounded, then G is a constant function where $\boldsymbol{C}$ denotes the set of complex numbers. That is to say, we need to choose either the regularity or the boundedness for an activation function of complex-valued neurons. In this paper, the expression (2) is adopted as an activation function of complex-valued neurons for the analysis, which is bounded but non-regular (that is, the boundedness is chosen).

There are some reasons for adopting the complex function defined in expression (2) as an activation function of complex-valued neurons, which we now state in the following.

First, although the expression (2) is not regular (i.e., the Cauchy-Riemann equation does not hold), there is a strong relationship between its real and imaginary parts. Actually, consider a complex-valued neuron with n-inputs, weights $w_k = w_k^r + iw_k^i \in \boldsymbol{C}$ $(1 \leq k \leq n)$, and a threshold value $\theta = \theta^r + i\theta^i \in \boldsymbol{C}$. Then, for n input signals $x_k + iy_k \in \boldsymbol{C}$ $(1 \leq k \leq n)$, the complex-valued neuron generates

$$\begin{aligned} X+iY &= f_C\Big(\sum_{k=1}^{n}(w_k^r+iw_k^i)(x_k+iy_k)+(\theta^r+i\theta^i)\Big) \\ &= f_R\Big(\sum_{k=1}^{n}(w_k^r x_k - w_k^i y_k)+\theta^r\Big) + if_R\Big(\sum_{k=1}^{n}(w_k^i x_k + w_k^r y_k)+\theta^i\Big) \end{aligned} \tag{3}$$

as an output. Both the real and imaginary parts of the right-hand side of the expression (3) contain the common variables $\{x_k, y_k, w_k^r, w_k^i\}_{k=1}^n$, and influence via those $4n$ variables each other. Moreover, the activation function (2) can process complex-valued signals properly through its amplitude-phase relationship. Actually, as described in (Nitta, 1997), the $1-n-1$ Complex-BP network with the activation function (2) can transform geometric figures, e.g. rotation, similarity transformation and parallel displacement of straight lines, circles, etc. This cannot be done without the ability to process signals properly through the amplitude-phase relationship in the activation function. Second, it has been proved that the complex-valued neural network with the activation function (2) can approximate any continuous complex-valued function, whereas the one with a regular activation function (for example, $f_C(z) = 1/(1+\exp(-z))$ proposed by Kim & Guest (1990), and $f_C(z) = \tanh(z)$ by Kim & Adali (2000)) cannot approximate any non-regular complex-valued function (Arena, Fortuna, Re, & Xibilia, 1993; Arena, Fortuna, Muscato, & Xibilia, 1998). That is, the complex-valued neural network with the activation function (2) is a universal approximator, and the one with a regular activation function is not, however, a universal approximator. Thus, the ability of complex-valued neural networks to approximate complex-valued functions depends heavily on the regularity of activation functions used. Third, the stability of the learning of the complex-valued neural network with the activation function (2) has been confirmed through some computer simulations (Nitta & Furuya, 1991; Nitta, 1993; Arena, Fortuna, Re, & Xibilia, 1993; Nitta, 1997). Finally, the activation function (2) clearly satisfies the following 5 properties an activation function $H(z) = u(x,y) + iv(x,y),\ z = x+iy$ should possess, which Georgiou & Koutsougeras (1992) pointed out:

1. H is nonlinear in x and y.
2. H is bounded.
3. The partial derivatives u_x, u_y, v_x and v_y exist and are bounded.
4. $H(z)$ is not entire. That is, there exists some $z_o \in \boldsymbol{C}$ such that $H(z_0)$ is not regular.
5. $u_x v_y$ is not identically equal to $v_x u_y$.

We adopted the activation function (2) for the above reasons.

3 Orthogonality of Decision Boundaries in the Complex-valued Neural Network

Decision boundary is a boundary by which the pattern classifier such as the Real-BP classifies patterns, and generally consists of hypersurfaces. Decision boundaries of real-valued neural networks have been examined empirically by Lippmann (1987). This section mathematically analyzes the decision boundaries of the complex-valued neural network.

3.1 A Case of a Single Neuron

We first analyze the decision boundary of a single complex-valued neuron (i.e., the number of hidden layers is zero).

Let the weights denote $\boldsymbol{w} = {}^t[w_1 \cdots w_n] = \boldsymbol{w}^r + i\boldsymbol{w}^i$, $\boldsymbol{w}^r = {}^t[w_1^r \cdots w_n^r]$, $\boldsymbol{w}^i = {}^t[w_1^i \cdots w_n^i]$, and let the threshold denote $\theta = \theta^r + i\theta^i$. Then, for n input signals (complex numbers) $\boldsymbol{z} = {}^t[z_1 \cdots z_n] = \boldsymbol{x} + i\boldsymbol{y}$, $\boldsymbol{x} = {}^t[x_1 \cdots x_n]$, $\boldsymbol{y} = {}^t[y_1 \cdots y_n]$, the complex-valued neuron generates

$$X + iY = f_R\left([{}^t\boldsymbol{w}^r \quad -{}^t\boldsymbol{w}^i]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} + \theta^r\right) + if_R\left([{}^t\boldsymbol{w}^i \quad {}^t\boldsymbol{w}^r]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} + \theta^i\right) \tag{4}$$

as an output. Here, for any two constants $C^R, C^I \in (0,1)$, let

$$X(\boldsymbol{x},\boldsymbol{y}) = f_R\left([{}^t\boldsymbol{w}^r \quad -{}^t\boldsymbol{w}^i]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} + \theta^r\right) = C^R, \tag{5}$$

$$Y(\boldsymbol{x},\boldsymbol{y}) = f_R\left([{}^t\boldsymbol{w}^i \quad {}^t\boldsymbol{w}^r]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} + \theta^i\right) = C^I. \tag{6}$$

Note here that expression (5) is the decision boundary for the *real part* of an output of the complex-valued neuron with n-inputs. That is, input signals $(\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{R}^{2n}$ are classified into two decision regions $\{(\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{R}^{2n} | X(\boldsymbol{x},\boldsymbol{y}) \geq C^R\}$ and $\{(\boldsymbol{x},\boldsymbol{y}) \in \boldsymbol{R}^{2n} | X(\boldsymbol{x},\boldsymbol{y}) < C^R\}$ by the hypersurface given by expression (5). Similarly, expression (6) is the decision boundary for the *imaginary part*. The normal vectors $Q^R(\boldsymbol{x},\boldsymbol{y})$ and $Q^I(\boldsymbol{x},\boldsymbol{y})$ of the decision boundaries ((5), (6)) are given by

$$\begin{aligned} Q^R(\boldsymbol{x},\boldsymbol{y}) &= \left(\frac{\partial X}{\partial x_1} \cdots \frac{\partial X}{\partial x_n} \frac{\partial X}{\partial y_1} \cdots \frac{\partial X}{\partial y_n}\right) \\ &= f_R'\left([{}^t\boldsymbol{w}^r \quad -{}^t\boldsymbol{w}^i]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} + \theta^r\right) \cdot [{}^t\boldsymbol{w}^r \quad -{}^t\boldsymbol{w}^i], \end{aligned} \tag{7}$$

$$\begin{aligned} Q^I(\boldsymbol{x},\boldsymbol{y}) &= \left(\frac{\partial Y}{\partial x_1} \cdots \frac{\partial Y}{\partial x_n} \frac{\partial Y}{\partial y_1} \cdots \frac{\partial Y}{\partial y_n}\right) \\ &= f_R'\left([{}^t\boldsymbol{w}^i \quad {}^t\boldsymbol{w}^r]\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} + \theta^i\right) \cdot [{}^t\boldsymbol{w}^i \quad {}^t\boldsymbol{w}^r]. \end{aligned} \tag{8}$$

Noting that the inner product of expressions (7) and (8) is zero, we can find that the decision boundary for the real part of an output of a complex-valued neuron and that for the imaginary part intersect orthogonally.

It can be easily shown that this orthogonal property also holds true for the other types of the complex-valued neurons proposed in (Kim & Guest, 1990; Benvenuto & Piazza, 1992; Georgiou & Koutsougeras, 1992). It should be noted here that there seems to be a problem on learning convergence in the formulation in (Kim & Guest, 1990); the complex-valued back-propagation algorithm with the activation function $f_C(z) = 1/(1+\exp(-z)), z = x+iy$ never converged in our experiments.

Generally, a real-valued neuron classifies an input real-valued signal into two classes (0, 1). On the other hand, a complex-valued neuron classifies an input complex-valued signal into four classes (0, 1, i, $1+i$). As described above, the decision boundary of a complex-valued neuron consists of two hypersurfaces which intersect orthogonally, and divides a decision region into four equal sections. Thus, it can be considered that a complex-valued neuron has a natural decision boundary for complex-valued patterns.

3.2 A Case of a Three-layered Network

Next, we examine the decision boundary of a three-layered complex-valued neural network (i.e., it has one hidden layer). Consider a three-layered complex-valued neural network with L input neurons, M hidden neurons, and N output neurons. We use $w_{ji} = w^r_{ji} + iw^i_{ji}$ for the weight between the input neuron i and the hidden neuron j, $v_{kj} = v^r_{kj} + iv^i_{kj}$ for the weight between the hidden neuron j and the output neuron k, $\theta_j = \theta^r_j + i\theta^i_j$ for the threshold of the hidden neuron j, $\gamma_k = \gamma^r_k + i\gamma^i_k$ for the threshold of the output neuron k. Then, for L input (complex-valued) signals $\boldsymbol{z} = {}^t[z_1 \cdots z_L] = \boldsymbol{x} + i\boldsymbol{y}$, $\boldsymbol{x} = {}^t[x_1 \cdots x_L]$, $\boldsymbol{y} = {}^t[y_1 \cdots y_L]$, the net input U_j to the hidden neuron j is given by

$$\begin{aligned} U_j &= U^r_j + iU^i_j \\ &= \Big[\sum_{i=1}^{L}(w^r_{ji}x_i - w^i_{ji}y_i) + \theta^r_j\Big] + i\Big[\sum_{i=1}^{L}(w^i_{ji}x_i + w^r_{ji}y_i) + \theta^i_j\Big]. \end{aligned} \tag{9}$$

Hence, the output H_j of the hidden neuron j is given by

$$H_j = H^r_j + iH^i_j = f_R(U^r_j) + if_R(U^i_j). \tag{10}$$

And also, the net input S_k to the output neuron k is given by

$$\begin{aligned} S_k &= S^r_k + iS^i_k \\ &= \Big[\sum_{j=1}^{M}(v^r_{kj}H^r_j - v^i_{kj}H^i_j) + \gamma^r_k\Big] + i\Big[\sum_{j=1}^{M}(v^i_{kj}H^r_j + v^r_{kj}H^i_j) + \gamma^i_k\Big]. \end{aligned} \tag{11}$$

Hence, the output O_k of the output neuron k is given by

$$O_k = O^r_k + iO^i_k = f_R(S^r_k) + if_R(S^i_k). \tag{12}$$

Here, for any two constants $C^R, C^I \in (0,1)$, let

$$O^r_k(\boldsymbol{x},\boldsymbol{y}) = C^R, \tag{13}$$

$$O^i_k(\boldsymbol{x},\boldsymbol{y}) = C^I. \tag{14}$$

The expressions (13) and (14) are the decision boundaries for the real and imaginary parts of the output neuron k in the 3-layered complex-valued neural network, respectively. The normal vectors $Q^R(\boldsymbol{x},\boldsymbol{y})$, $Q^I(\boldsymbol{x},\boldsymbol{y})$ of these hypersurfaces ((13), (14)) are given by

$$Q^R(\boldsymbol{x},\boldsymbol{y}) = \Big(\frac{\partial O_k^r}{\partial x_1} \ \cdots \ \frac{\partial O_k^r}{\partial x_L} \ \frac{\partial O_k^r}{\partial y_1} \ \cdots \ \frac{\partial O_k^r}{\partial y_L}\Big), \tag{15}$$

$$Q^I(\boldsymbol{x},\boldsymbol{y}) = \Big(\frac{\partial O_k^i}{\partial x_1} \ \cdots \ \frac{\partial O_k^i}{\partial x_L} \ \frac{\partial O_k^i}{\partial y_1} \ \cdots \ \frac{\partial O_k^i}{\partial y_L}\Big), \tag{16}$$

and their inner product is given by

$$\begin{aligned} & Q^R(\boldsymbol{x},\boldsymbol{y}) \cdot {}^tQ^I(\boldsymbol{x},\boldsymbol{y}) \\ = \ & \frac{\partial O_k^r}{\partial x_1}\cdot\frac{\partial O_k^i}{\partial x_1} + \ \cdots \ + \frac{\partial O_k^r}{\partial x_L}\cdot\frac{\partial O_k^i}{\partial x_L} + \frac{\partial O_k^r}{\partial y_1}\cdot\frac{\partial O_k^i}{\partial y_1} + \ \cdots \ + \frac{\partial O_k^r}{\partial y_L}\cdot\frac{\partial O_k^i}{\partial y_L}. \end{aligned} \tag{17}$$

Note here that, for any $1 \le i \le L$,

$$\begin{aligned} & \frac{\partial O_k^r}{\partial x_i}\cdot\frac{\partial O_k^i}{\partial x_i} + \frac{\partial O_k^r}{\partial y_i}\cdot\frac{\partial O_k^i}{\partial y_i} \\ = \ & \frac{\partial f_R(S_k^r)}{\partial S_k^r}\cdot\frac{\partial f_R(S_k^i)}{\partial S_k^i}\cdot\Big[\sum_{j=1}^M\Big(v_{kj}^r w_{ji}^r\cdot\frac{\partial f_R(U_j^r)}{\partial U_j^r} - v_{kj}^i w_{ji}^i\cdot\frac{\partial f_R(U_j^i)}{\partial U_j^i}\Big)\Big] \\ & \qquad \cdot\Big[\sum_{j=1}^M\Big(v_{kj}^i w_{ji}^r\cdot\frac{\partial f_R(U_j^r)}{\partial U_j^r} + v_{kj}^r w_{ji}^i\cdot\frac{\partial f_R(U_j^i)}{\partial U_j^i}\Big)\Big] \\ & -\frac{\partial f_R(S_k^r)}{\partial S_k^r}\cdot\frac{\partial f_R(S_k^i)}{\partial S_k^i}\cdot\Big[\sum_{j=1}^M\Big(v_{kj}^r w_{ji}^r\cdot\frac{\partial f_R(U_j^i)}{\partial U_j^i} - v_{kj}^i w_{ji}^i\cdot\frac{\partial f_R(U_j^r)}{\partial U_j^r}\Big)\Big] \\ & \qquad \cdot\Big[\sum_{j=1}^M\Big(v_{kj}^i w_{ji}^r\cdot\frac{\partial f_R(U_j^i)}{\partial U_j^i} + v_{kj}^r w_{ji}^i\cdot\frac{\partial f_R(U_j^r)}{\partial U_j^r}\Big)\Big]. \end{aligned} \tag{18}$$

Hence, the inner product of the normal vectors is not always zero. Therefore, we can not conclude that the decision boundaries (hypersurfaces) for the real and imaginary parts of the output neuron k in the 3-layered complex-valued neural network intersect orthogonally. However, paying enough attention to expression (18), we can find that if

$$\frac{\partial f_R(U_j^r)}{\partial U_j^r} = \frac{\partial f_R(U_j^i)}{\partial U_j^i} \tag{19}$$

for any $1 \le j \le M$, then the inner product is zero. In general, if both $|u_1|$ and $|u_2|$ are sufficiently large, we can consider that $f_R'(u_1)$ is nearly equal to $f_R'(u_2)$. Hence, if, for any $1 \le j \le M$, there exist sufficiently large positive real numbers K_1, K_2 such that $|U_j^r| > K_1$ and $|U_j^i| > K_2$, then the two decision boundaries ((13), (14)) nearly intersect orthogonally. That is, if, for any $1 \le j \le M$, both the absolute values of the real and imaginary parts of the net input (complex number) to the hidden neuron j are sufficiently large, then the decision boundaries nearly intersect orthogonally. Therefore, the following theorem can

be obtained.

Theorem The decision boundaries for the real and imaginary parts of an output neuron in the 3-layered complex-valued neural network approach orthogonality as both the absolute values of the real and imaginary parts of the net inputs to all hidden neurons grow. □

4 Utility of the Orthogonal Decision Boundaries

In this section, we will show the utility which the properties on the decision boundary discovered in the previous section bring about.

Minsky and Papert (1969) clarified the limitations of 2-layered real-valued neural networks (i.e., no hidden layers): in a large number of interesting cases, the 2-layered real-valued neural network is incapable of solving the problems. A classic example of this case is the exclusive-or (XOR) problem which has a long history in the study of neural networks, and many other difficult problems involve the XOR as subproblem. Another example is the detection of symmetry problem. Rumelhart, Hinton, & Williams (1986a, 1986b) showed that the *3-layered* real-valued neural network (i.e., with one hidden layer) can solve such problems including the XOR problem and the detection of symmetry problem and the interesting internal representations can be constructed in the weight-space.

As described above, the XOR problem and the detection of symmetry problem cannot be solved with the 2-layered real-valued neural network. Then, first, contrary to expectation, it will be proved that such problems can be solved by the *2-layered* complex-valued neural network (i.e., no hidden layers) with the orthogonal decision boundaries, which reveals a potent computational power of complex-valued neural nets. In addition, it will be shown as an application of the above computational power that the fading equalization problem can be successfully solved by the 2-layered complex-valued neural network with the highest generalization ability. Rumelhart, Hinton and Williams (1986a, 1986b) showed that increasing the number of layers made the computational power of neural networks high. In this section, we will show that extending the dimensionality of neural networks to complex numbers originates the similar effect on neural networks. This may be a new directionality for enhancing the ability of neural networks.

Second, we will present the simulation results on the generalization ability of the *3-layered* complex-valued neural networks trained using the Complex-BP (called *Complex-BP network*) (Nitta & Furuya, 1991; Nitta, 1993, 1997) and will compare them with those of the 3-layered real-valued neural networks trained using the Real-BP (called *Real-BP network*) (Rumelhart, Hinton, & Williams, 1986a, 1986b).

4.1 The XOR Problem

In this section, it is proved that the XOR problem can be solved by *2-layered* complex-valued neural network (i.e., no hidden layers) with the orthogonal decision boundaries.

The input-output mapping in the XOR problem is shown in Table 1.

Table 1 The XOR problem

Input		Output
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

In order to solve the XOR problem with complex-valued neural networks, the input-output mapping is encoded as shown in Table 2 where the outputs 1 and i are interpreted to be 0, 0 and $1+i$ are interpreted to be 1 of the original XOR problem (Table 1), respectively.

Table 2 An encoded XOR problem for complex-valued neural networks

Input	Output
$z = x + iy$	$Z = X + iY$
$-1-i$	1
$-1+i$	0
$1-i$	$1+i$
$1+i$	i

We use a 1-1 complex-valued neural network (i.e., no hidden layers) with a weight $w = u + iv \in \boldsymbol{C}$ between the input neuron and the output neuron (we assume that it has no threshold parameters). The activation function is defined to be

$$1_C(z) = 1_R(x) + i1_R(y), \quad z = x + iy \tag{20}$$

where 1_R is a real-valued step function defined on $\boldsymbol{R}$, that is, $1_R(u) = 1$ if $u \geq 0$, 0 otherwise for any $u \in \boldsymbol{R}$. The decision boundary of the 1-1 complex-valued neural network described above consists of the following two straight lines which intersect orthogonally:

$$[u \quad -v] \cdot {}^t[x \quad y] = 0, \tag{21}$$

$$[v \quad u] \cdot {}^t[x \quad y] = 0 \tag{22}$$

for any input signal $z = x + iy \in \boldsymbol{C}$ where u and v are the real and imaginary parts of the weight parameter $w = u + iv$, respectively. The expressions (21) and (22) are the decision boundaries for the real and imaginary parts of the 1-1 complex-valued neural network, respectively. Letting $u = 0$ and $v = 1$ (i.e., $w = i$), we have the decision boundary shown

in Fig. 1, which divides the input space (the decision region) into four equal sections, and has the highest generalization ability for the XOR problem. On the other hand, the decision boundary of the 3-layered real-valued neural network for the XOR problem does not always have the highest generalization ability (Lippmann, 1987). In addition, the required number of learnable parameters is only 2 (i.e., only $w = u + iv$), whereas at least 9 parameters are needed for the 3-layered real-valued neural network to solve the XOR problem (Rumelhart, Hinton, & Williams, 1986a, 1986b), where a complex-valued parameter $z = x + iy$ (where $i = \sqrt{-1}$) is counted as two because it consists of a real part x and an imaginary part y.

4.2 The Detection of Symmetry

Another interesting task that cannot be done by 2-layered real-valued neural networks is the detection of symmetry problem (Minsky and Papert, 1969). In this section, a solution to this problem using *2-layered* complex-valued neural network (i.e., no hidden layers) with the orthogonal decision boundaries is given.

The problem is to detect whether the binary activity levels of a one-dimensional array of input neurons are symmetrical about the centre point. For example, the input-output mapping in the case of 3 inputs is shown in Table 3. We used patterns of various lengths (from 2 to 6) and could solve all the cases with 2-layered complex-valued neural networks. Only a solution to the case with 6 inputs is presented here because the other cases can be done by the similar way.

Table 3 The detection of symmetry problem with 3 inputs. Output 1 means that the corresponding input is symmetric, and 0 asymmetric.

Input			Output
x_1	x_2	x_3	y
0	0	0	1
0	0	1	0
0	1	0	1
1	0	0	0
0	1	1	0
1	0	1	1
1	1	0	0
1	1	1	1

We use a 6-1 complex-valued neural network (i.e., no hidden layers) with weights $w_k = u_k + iv_k \in \boldsymbol{C}$ between an input neuron k and the output neuron ($1 \leq k \leq 6$) (we assume that it has no threshold parameters). In order to solve the problem with the complex-valued neural network, the input-output mapping is encoded as follows: an input $x_k \in \boldsymbol{R}$ is encoded as an input $x_k + iy_k \in \boldsymbol{C}$ to the input neuron k where $y_k = 0$ ($1 \leq k \leq 6$), the output $1 \in \boldsymbol{R}$ is encoded as $1 + i \in \boldsymbol{C}$, and the output $0 \in \boldsymbol{R}$ is

encoded as 1 or $i \in \boldsymbol{C}$ which is determined according to inputs (for example, the output corresponding to the input ${}^t[0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0]$ is i). The activation function is the same as in expression (20).

The decision boundary of the 6-1 complex-valued neural network described above consists of the following two straight lines which intersect orthogonally:

$$[u_1\ \ \cdots\ \ u_6\ \ -v_1\ \ \cdots\ \ -v_6] \cdot {}^t[x_1\ \ \cdots\ \ x_6\ \ y_1\ \ \cdots\ \ y_6] = 0, \tag{23}$$

$$[v_1\ \ \cdots\ \ v_6\ \ u_1\ \ \cdots\ \ u_6] \cdot {}^t[x_1\ \ \cdots\ \ x_6\ \ y_1\ \ \cdots\ \ y_6] = 0 \tag{24}$$

for any input signal $z_k = x_k + iy_k \in \boldsymbol{C}$ where u_k and v_k are the real and imaginary parts of the weight parameter $w_k = u_k + iv_k$, respectively ($1 \leq k \leq 6$). The expressions (23) and (24) are the decision boundaries for the real and imaginary parts of the 6-1 complex-valued neural network, respectively. Letting ${}^t[u_1\ \cdots\ u_6] = {}^t[-1\ \ 2\ \ -4\ \ 4\ \ -2\ \ 1]$ and ${}^t[v_1\ \cdots\ v_6] = {}^t[1\ \ -2\ \ 4\ \ -4\ \ 2\ \ -1]$ (i.e., $w_1 = -1 + i, w_2 = 2 - 2i, w_3 = -4 + 4i, w_4 = 4 - 4i, w_5 = -2 + 2i$ and $w_6 = 1 - i$), we have the orthogonal decision boundaries shown in Fig. 2 which successfully detect the symmetry of the $2^6 (= 64)$ input patterns.

In addition, the required number of learnable parameters is 12 (i.e., 6 complex-valued weights), whereas at least 17 parameters are needed for the 3-layered real-valued neural network to solve the detection of symmetry (Rumelhart, Hinton, & Williams, 1986a, 1986b) where a complex-valued parameter $z = x + iy$ (where $i = \sqrt{-1}$) is counted as two as in Section 4.1.

4.3 The Fading Equalization Technology

In this section, it is shown that 2-layered complex-valued neural networks with orthogonal decision boundaries can be successfully applied to the fading equalization technology (Lathi, 1998).

Channel equalization in a digital communication system can be viewed as a pattern classification problem. The digital communication system receives a transmitted signal sequence with additive noise, and tries to estimate the true transmitted sequence. A transmitted signal can take one of the following four possible complex values: $-1 - i, -1 + i, 1 - i$ and $1 + i$ ($i = \sqrt{-1}$). Thus the received signal will take value around $-1 - i, -1 + i, 1 - i$ and $1 + i$, for example, $-0.9 - 1.2i$, $1.1 + 0.84i$ or something because some noises are added to them. We need to estimate the true complex values from such complex values with noises. Thus, the method with an excellent generalization ability is needed for the estimate. The input-output mapping in the problem is shown in Table 4.

Table 4 The input-output mapping in the fading equalization problem

Input	Output
$-1 - i$	$-1 - i$
$-1 + i$	$-1 + i$
$1 - i$	$1 - i$
$1 + i$	$1 + i$

We use the same 1-1 complex-valued neural network as in Section 4.1. In order to solve the problem with the complex-valued neural network, the input-output mapping in Table 4 is encoded as shown in Table 5. Letting $u = 1$ and $v = 0$ (i.e., $w = 1$), we have the orthogonal decision boundary shown in Fig. 3, which has the highest generalization ability for the fading equalization problem, and can estimate true signals without errors. In addition, the required number of learnable parameters is only 2 (i.e., only $w = u + iv$).

Table 5 An encoded fading equalization problem for complex-valued neural networks

Input	Output
$-1-i$	0
$-1+i$	i
$1-i$	1
$1+i$	$1+i$

4.4 Generalization Ability of 3-layered Complex-valued Neural Networks

We present below the simulation results on the generalization ability of the three-layered complex-valued neural networks trained using the Complex-BP (called *Complex-BP network*) (Nitta & Furuya, 1991; Nitta, 1993, 1997) and compare them with those of the three-layered real-valued neural networks trained using the Real-BP (called *Real-BP network*) (Rumelhart, Hinton, & Williams, 1986a, 1986b).

In the experiments, the three sets of (complex-valued) learning patterns shown in Tables 6-8 were used, and the learning constant ε was 0.5. The initial components of the weights and the thresholds were chosen to be random real numbers between -0.3 and 0.3. We judged that learning finished, when $\sqrt{\sum_p \sum_{k=1}^{N} |T_k^{(p)} - O_k^{(p)}|^2} = 0.05$ held, where $T_k^{(p)}$, $O_k^{(p)} \in \boldsymbol{C}$ denoted the desired output value, the actual output value of the output neuron k for the pattern p; N denoted the number of neurons in the output layer. We regarded presenting a set of learning patterns to the neural network as one learning cycle.

Table 6 Learning pattern 1

Input pattern	*Output pattern*
$-0.03-0.03i$	$1+i$
$0.03-0.03i$	i
$0.03+0.03i$	0
$-0.03+0.03i$	1

Table 7 Learning pattern 2

Input pattern	*Output pattern*
$-0.03-0.03i$	i
$0.03-0.03i$	0
$0.03+0.03i$	1
$-0.03+0.03i$	$1+i$

Table 8 Learning pattern 3

Input pattern	*Output pattern*
$-0.03-0.03i$	0
$0.03-0.03i$	1
$0.03+0.03i$	$1+i$
$-0.03+0.03i$	i

We used the four kinds of three-layered Complex-BP networks: 1-3-1, 1-6-1, 1-9-1 and 1-12-1 networks. After training, by presenting the 1,681(=41×41) points in the complex plane $[-1,1]\times[-1,1]$ ($x+iy$, where $x=-1.0,-0.95,\cdots,0.95,1.0$; $y=-1.0,-0.95,\cdots,0.95$, 1.0), the actual output points formed the decision boundaries. Fig. 4 shows an example of the decision boundary of the Complex-BP network. In Fig. 4, the number **1** denotes the region in which the real part of the output value of the neural network is **OFF** (0.0-0.5), and the imaginary part **OFF**; the region **2** the real part **ON** (0.5-1.0), and the imaginary part **OFF**; the region **3** the real part **OFF**, and the imaginary part **ON**; the region **4** the real part **ON**, and the imaginary part **ON**. And the decision boundary for the real part (i.e., the boundary that the region "**1+3**" and the region "**2+4**" form) and that for imaginary part (i.e., the boundary that the region "**1+2**" and the region "**3+4**" form) intersect orthogonally.

We also conducted the corresponding experiments for the Real-BP networks. We chose the 2-4-2 Real-BP network for the 1-3-1 Complex-BP network as a comparison object because the numbers of the parameters (weights and thresholds) were almost the same: the number of parameters for the 1-3-1 Complex-BP network was 20, and that for the 2-4-2 Real-BP network 22 where a complex-valued parameter $z=x+iy$ (where $i=\sqrt{-1}$) was counted as two because it consisted of a real part x and an imaginary part y. Similarly, the 2-7-2, 2-11-2 and 2-14-2 Real-BP networks were chosen for the 1-6-1, 1-9-1 and 1-12-1 Complex-BP networks as their comparison objects, respectively. The numbers of parameters of them are shown in Table 9. In the Real-BP networks, the real component of a complex number was input into the first input neuron, and the imaginary component was input into the second input neuron; the output from the first output neuron was interpreted to be the real component of a complex number, and the output from the second output neuron was interpreted to be the imaginary component. Fig. 5 shows an example of the decision boundary of the Real-BP network where the numbers **1-4** have the same meanings as those of Fig. 4. We can find from Fig. 5 that the decision boundary for the real part (i.e., the boundary that the region "**1+3**" and the region "**2+4**" form) and that for imaginary part (i.e., the boundary that the region "**1+2**" and the region "**3+4**" form) do not intersect orthogonally.

Table 9 The number of parameters in the Real-BP and Complex-BP networks

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
The number of parameters	20	38	56	74
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
The number of parameters	22	37	57	72

First, we measured the angles between the decision boundary for the real part (i.e., the boundary that the region "**1+3**" and the region "**2+4**" formed) and that for imaginary part (i.e., the boundary that the region "**1+2**" and the region "**3+4**" formed) which were the components of the decision boundary of the output neuron in the Complex-BP networks in the visual observation under the experimental conditions described above. In the visual observation, the angles of the decision boundaries observed were roughly classified into three classes: 30, 60 and 90 degrees. The result of the observation is shown in Table 10 where most of trials were 90 degrees, few trials 60 degrees and none of trials 30 degrees. Then, the average and the standard deviation of the angles of 100 trials for each of the 3 learning patterns and each of the 4 kinds of network structures were used as the evaluation criterion. Although we stopped learning at the 200,000th iteration, all trials succeeded in converging. And also, we measured the same quantities of the Real-BP networks for the comparison. The results of the experiments are shown in Table 11. We can find from Table 11 that all the average angles for the Complex-BP networks are almost 90 degrees, which are independent of the learning patterns and the network structures, whereas those of the Real-BP networks are around 70-80 degrees. In addition, the standard deviations of the angles for the Complex-BP networks are around 0-5 degrees and those for the Real-BP networks around 20 degrees. Thus, we can conclude from the experimental results that the decision boundary for the real part and that for imaginary part which are the components of the decision boundary of the output neuron in the three-layered Complex-BP networks almost intersect orthogonally, whereas those for the Real-BP networks do not.

Table 10 Result of the visual observation of the angles of the decision boundaries (Complex-BP network). Angles were roughly classified into the three classes: 30, 60 and 90 degrees. A numeral means the number of trials in which the angle observed was classified into a specific class.

	30 degrees	60 degrees	90 degrees
Learning pattern 2 (1-3-1 network)	0	4	96
Learning pattern 3 (1-6-1 network)	0	3	97
Learning pattern 3 (1-9-1 network)	0	1	99
Other 9 cases	0	0	100

Table 11. Comparison of the angles of the decision boundaries (the average and the standard deviation). The unit is degree.

(a) Pattern 1

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	90	90	90	90
Standard deviation	0	0	0	0
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	78	72	76	80
Standard deviation	19	22	17	17

(b) Pattern 2

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	89	90	90	90
Standard deviation	6	0	0	0
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	85	77	77	75
Standard deviation	16	20	18	21

(c) Pattern 3

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	90	89	90	90
Standard deviation	0	5	3	0
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	86	76	72	73
Standard deviation	11	20	22	22

It seems that there is a possibility of the improvement of the generalization ability of neural networks caused by the orthogonality of the decision boundaries of the network. Next, we measured the discrimination rate of the Complex-BP network for unlearned patterns in order to clarify how the orthogonality of the decision boundary of the 3-layered Complex-BP network changed its generalization ability. To specifically, we counted the number of the test patterns for which the Complex-BP network could give the correct output in the same experiments described above on the angles of decision boundaries of 100 trials for each of the 3 learning patterns and each of the 4 kinds of network structures. We defined the correctness as follows: the output value $X + iY (0 \leq X, Y \leq 1)$ of the Complex-BP network for an unlearned pattern $x + iy$ $(-1 \leq x, y \leq 1)$ was correct if $|X - A| < 0.5$ and $|Y - B| < 0.5$, provided that the closest input learning pattern to the unlearned pattern $x + iy$ was $a + ib$ whose corresponding output learning pattern was $A + iB$ $(A, B = 0$ or $1)$. For example, the output value $X + iY (0 \leq X, Y \leq 1)$ of the Complex-BP network for an unlearned pattern $x + iy$ $(0 \leq x, y \leq 1)$ was correct if both the real and imaginary parts of the output value of the Complex-BP network took value less than 0.5, provided that the corresponding output learning pattern for the input

learning pattern $0.03 + 0.03i$ was 0. Then, the average and the standard deviation of the discrimination rate of 100 trials for each of the 3 learning patterns and each of the 4 kinds of network structures were used as the evaluation criterion. The results of the experiments including the Real-BP network case appear in Table 12. The above simulation results clearly suggest that the Complex-BP network has better generalization performance than that of the Real-BP network. Furthermore, we investigated the causality between the high generalization ability of the Complex-BP network and the orthogonal decision boundary. Table 13 shows the average of the discrimination rate for the three upper cases shown in Table 10. It is clearly suggested from Table 13 that the average of the discrimination rate for the 3-layered complex-valued neural network with the orthogonal decision boundary is superior to that for the one with the non-orthogonal decision boundary. Thus, we believe that the orthogonality of the decision boundaries causes the high generalization ability of the Complex-BP network.

Table 12. Comparison of the discrimination rate (the average and the standard deviation). The unit is percentage (%).

(a) Pattern 1

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	92	95	97	98
Standard deviation	6	5	3	2
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	88	90	93	93
Standard deviation	8	7	4	4

(b) Pattern 2

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	93	95	96	97
Standard deviation	6	4	4	3
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	88	91	92	93
Standard deviation	8	7	5	6

(c) Pattern 3

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	92	94	97	97
Standard deviation	7	4	3	3
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	87	90	90	92
Standard deviation	9	7	7	6

Table 13 The average of the discrimination rate for the three upper cases shown in Table 10. The unit is percentage (%).

	30 degrees	60 degrees	90 degrees
Learning pattern 2 (1-3-1 network)	—	88	93
Learning pattern 3 (1-6-1 network)	—	88	93
Learning pattern 3 (1-9-1 network)	—	84	96

Finally, we investigated the average and the standard deviation of the learning speed (i.e., learning cycles needed to converge) of 100 trials for each of the 3 learning patterns and each of the 4 kinds of network structures in the experiments described above. The results of the experiments are shown in Table 14. We can find from these experiments that the learning speed of the Complex-BP is several times faster than that of the Real-BP, and the standard deviation of the learning speed of the Complex-BP is smaller than that of the Real-BP.

Table 14. Comparison of the learning speed (the average and the standard deviation). The unit is learning cycle.

(a) Pattern 1

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	10770	10178	9766	9529
Standard deviation	438	472	210	144
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	31647	29945	28947	28230
Standard deviation	1268	944	697	566

(b) Pattern 2

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	10608	9932	9713	9539
Standard deviation	418	167	148	110
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	31781	29842	28902	28267
Standard deviation	2126	809	721	576

(c) Pattern 3

Complex-BP network	1-3-1	1-6-1	1-9-1	1-12-1
Average	10746	10055	9713	9502
Standard deviation	740	412	228	188
Real-BP network	2-4-2	2-7-2	2-11-2	2-14-2
Average	34038	29620	28980	28471
Standard deviation	5182	806	603	584

5 Discussion

We have proved that the decision boundary for the real part of an output of a single complex-valued neuron and that for the imaginary part intersect orthogonally in Section 3.1. Since this property is completely different from an usual real-valued neuron, one needs to design the complex-valued neural network for real applications and its learning algorithm taking into account the orthogonal property of the complex-valued neuron whatever the type of the network is (multi-layered type or mutually-connected type). Moreover, it has constructively been proved using the orthogonal property of the decision boundary that the XOR problem and the detection of symmetry problem can be solved by 2-layered complex-valued neural networks (i.e., complex-valued neurons), which cannot be solved by 2-layered real-valued neural networks (i.e., real-valued neurons). These results reveal a potent computational power of complex-valued neural nets. Making the dimensionality of neural networks high (for example, complex numbers) may be a new directionality for enhancing the ability of neural networks. In addition, it has been shown as an application of the above computational power that the fading equalization problem can be successfully solved by the 2-layered complex-valued neural network with the highest generalization ability. One should use 2-layered complex-valued neural networks rather than 3-layered ones when solving the fading equalization problem with complex-valued neural networks.

It is not always guaranteed that the decision boundary of the 3-layered complex-valued neural network has the orthogonality, as we have made clear in Section 3.2. Then, we have derived a sufficient condition for the decision boundaries in the 3-layered complex-valued neural network to almost intersect orthogonally in Section 3.2 (Theorem). The sufficient condition was as follows: both the absolute values of the real and imaginary parts of the net inputs to all hidden neurons are sufficiently large. This is a *characterization* for the structure of the decision boundaries in the 3-layered complex-valued neural network. The theorem will be useful if a learning algorithm such that both the absolute values of the real and imaginary parts of the net inputs to all hidden neurons become sufficiently large is devised because there is a possibility that the orthogonality of the decision boundaries of the network can improve the generalization ability of 3-layered complex-valued neural networks as we have seen in Section 4.4. That is, there is a possibility that we can utilize the theorem in order to improve the generalization ability of the 3-layered complex-valued neural network. However, the situation in which the theorem is directly useful for the Complex-BP network cannot be considered regrettably for now because the control of the net input is difficult as long as the Complex-BP algorithm (that is, steepest descent method) is used. The Complex-BP is one of the learning algorithms for complex-valued neural networks. Thus it should be noted that the usefulness of the theorem depends on the learning algorithm used. Although the orthogonality of the decision boundaries in the 3-layered complex-valued neural network can be guaranteed conditionally as described above, we can find from the experiments in Section 4.4 that most of the decision boundaries in the 3-layered Complex-BP network intersect orthogonally. Moreover, it is learned from the experiments that there is a possibility that the orthogonality of the decision boundaries in the 3-layered Complex-BP network improves its generalization

ability. The decision boundary of the complex-valued neural network which consists of two orthogonal hypersurfaces divides a decision region into four equal sections. So, it is intuitively considered that the orthogonality of the decision boundaries improves its generalization ability. Then, we showed the possibility experimentally. The theoretical evaluation of the generalization ability of the complex-valued neural network using the criteria for the evaluation such as the Vapnik-Chervonenkis dimension (VC-dimension) (Vapnik, 1998) could clarify how the orthogonal property of the decision boundaries of the complex-valued neural network influences its generalization ability.

It had already been reported that the average of the learning speed of the Complex-BP is several times faster than that of the Real-BP (Nitta, 1997). In this connection, we could confirm this again in the experiments on the orthogonality of the decision boundary and the generalization ability of the 3-layered Complex-BP network in Section 4.4. It was learned that the standard deviation of the learning speed of the Complex-BP was smaller than that of the Real-BP, which had not been reported in (Nitta, 1997).

6 Conclusions

We have made clear the differences between the real-valued neural network and the complex-valued neural network through theoretical and experimental analyses of their fundamental properties, and clarified the utility for the complex-valued neural network which the properties discovered in this paper bring about.

It turned out that the complex-valued neural network has some inherent properties on decision boundary, as a result of the extension to complex numbers. A decision boundary of a single complex-valued neuron consists of two hypersurfaces which intersect orthogonally, and divides a decision region into four equal sections. The XOR problem and the detection of symmetry problem which cannot be solved with 2-layered real-valued neural networks, can be solved by 2-layered complex-valued neural networks with the orthogonal decision boundaries, which reveals a potent computational power of complex-valued neural nets. Furthermore, the fading equalization problem can be successfully solved by the 2-layered complex-valued neural network with the highest generalization ability. A decision boundary of a three-layered complex-valued neural network has the orthogonal property as a basic structure, and its two hypersurfaces approach orthogonality as all the net inputs to each hidden neuron grow. In particular, most of the decision boundaries in the 3-layered complex-valued neural network intersect orthogonally when the network is trained using the Complex-BP algorithm. As a result, the orthogonality of the decision boundaries improves its generalization ability. Its theoretical proof is a future topic. Furthermore, the average of the learning speed of the Complex-BP is several times faster than that of the Real-BP. The standard deviation of the learning speed of the Complex-BP is smaller than that of the Real-BP. The complex-valued neural network and the related Complex-BP algorithm are the natural methods to learn complex-valued patterns for the above reasons, and are expected to be effectively used in the fields dealing with complex numbers.

Acknowledgements
The author would like to give special thanks to Drs. S.Akaho, Y.Akiyama, M.Asogawa, T.Furuya and the anonymous reviewers for valuable comments.

References

Arena, P., Fortuna, L., Re, R., & Xibilia, M. G. (1993). On the capability of neural networks with complex neurons in complex valued functions approximation. In *Proc. IEEE Int. Conf. on Circuits and Systems* (pp.2168-2171).

Arena, P., Fortuna, L., Muscato, G., & Xibilia, M. G. (1998). *Neural networks in multi-dimensional domains.* Lecture Notes in Control and Information Sciences **234**, Springer.

Benvenuto, N., & Piazza, F. (1992). On the complex backpropagation algorithm. *IEEE Trans. Signal Processing*, 40(4), 967-969.

Georgiou, G. M., & Koutsougeras, C. (1992). Complex domain backpropagation. *IEEE Trans. Circuits and Systems–II: Analog and Digital Signal Processing*, 39(5), 330-334.

ICONIP. (2002). 6 papers in the special session: Complex-valued neural networks. In *Proc. International Conference on Neural Information Processing*, Singapore (Vol.3, pp.1074-1103).

KES. (2001). 6 papers in the special session: Complex-valued neural networks and their applications. In N. Baba, L. C. Jain, & R. J. Howlett (Eds.), *Knowledge-based Intelligent Information Engineering Systems & Allied Technologies*, Part I (pp.550-580). IOS Press, Tokyo.

KES. (2002). 5 papers in the special session: Complex-valued neural networks. In E. Damiani, R. J. Howlett, L. C. Jain, & N. Ichalkaranje (Eds.), *Knowledge-based Intelligent Information Engineering Systems & Allied Technologies*, Part I (pp.623-647). IOS Press, Amsterdam.

Kim, T., & Adali, T. (2000). Fully complex backpropagation for constant envelope signal processing. In *Proc. IEEE Workshop on Neural Networks for Signal Processing* (pp.231-240).

Kim, M. S., & Guest, C. C. (1990). Modification of backpropagation networks for complex-valued signal processing in frequency domain. In *Proc. Int. Joint Conf. on Neural Networks* (Vol.3, pp.27-31).

Kuroe, Y., Hashimoto, N., & Mori, T. (2002). On energy function for complex-valued neural networks and its applications. In *Proc. Int. Conf. on Neural Information Processing* (Vol.3, pp.1079-1083).

Lathi, B. P. (1998). *Modern digital and analog communication systems.* Third Edition Oxford University Press -USA.

Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE Acoustic, Speech and Signal Processing Magazine*, April, 4-22.

Minsky, M. L., & Papert, S. A. (1969). *Perceptrons.* MIT Press, Cambridge.

Miyauchi, M., Seki, M., Watanabe, A., & Miyauchi, A. (1992). Interpretation of optical flow through neural network learning. In *Proc. IAPR Workshop on Machine Vision Applications* (pp.523-528).

Miyauchi, M., Seki, M., Watanabe, A., & Miyauchi, A. (1993). Interpretation of optical flow through complex neural network. In *Proc. Int. Workshop on Artificial Neural Networks: Lecture Notes in Computer Science* (686, pp.645-650), Springer-Verlag.

Nitta, T., & Furuya, T. (1991). A complex back-propagation learning. *Transactions of Information Processing Society of Japan*, 32(10), 1319-1329 (in Japanese).

Nitta, T. (1993). A complex numbered version of the back-propagation algorithm. In *Proc. World Congress on Neural Networks* (Vol.3, pp.576-579).

Nitta, T. (1997). An extension of the back-propagation algorithm to complex numbers. *Neural Networks*, 10(8), 1392-1415.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986a). *Parallel distributed processing* (Vol.1). MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986b). Learning representations by back-propagating errors. *Nature*, 323, 533-536.

Vapnik, V. N. (1998). *Statistical learning theory.* John Wiley & Sons, Inc.

Watanabe, A., Yazawa, N., Miyauchi, A., & Miyauchi, M. (1994). A method to interpret 3D motions using neural networks. *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, E77-A(8), 1363-1370.

Figure 1: The decision boundary in the input space of the 1-1 complex-valued neural network that solves the XOR problem. The black circle means that the output in the XOR problem is 1, and the white one 0.

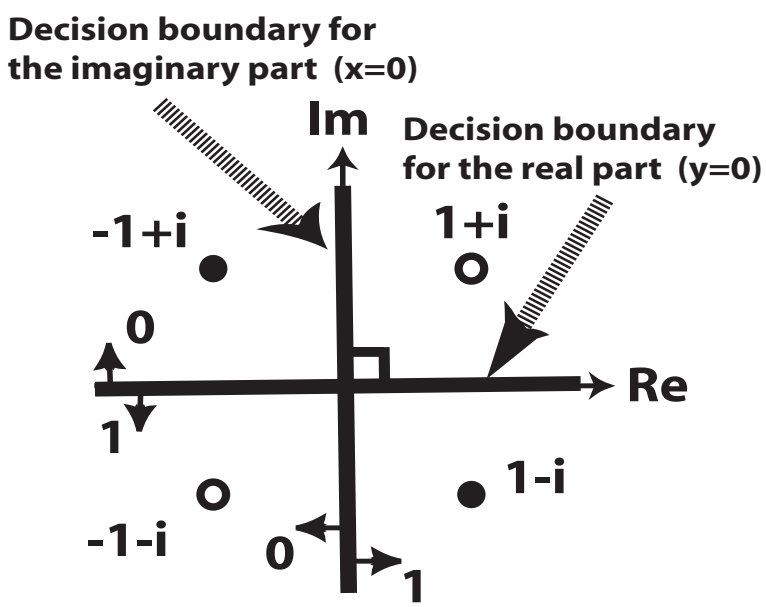

Figure 2: The decision boundary in the net-input space of the 6-1 complex-valued neural network that solves the detection of symmetry problem. Note that the plane is not the *input* space but the *net-input* space because the dimension of the input space is 6 and the input space cannot be written in a 2-dimensional plane. The black circle means a net-input for a symmetric input, and the white one asymmetric. There is only one black circle at the origin. The four circled complex numbers mean the output values of the 6-1 complex-valued neural network in their regions, respectively.

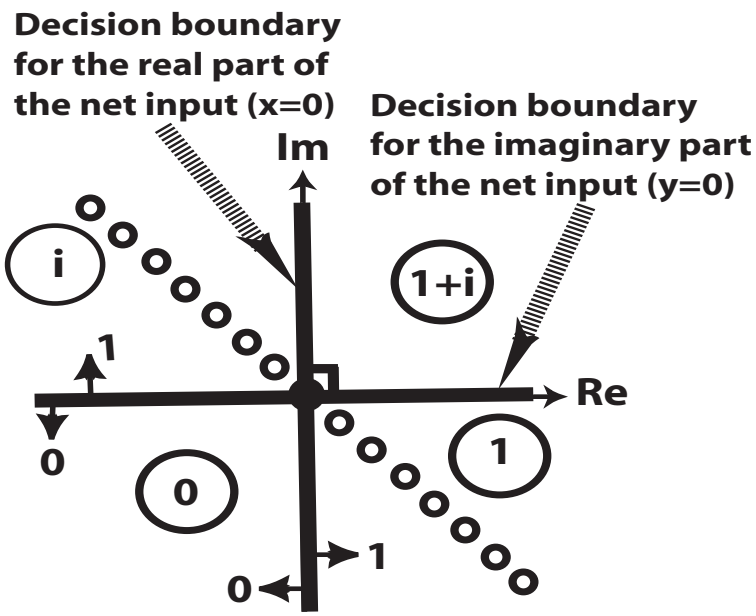

Figure 3: The decision boundary in the input space of the 1-1 complex-valued neural network that solves the fading equalization problem. The black circle means an input in the fading equalization problem. The four circled complex numbers mean the output values of the 1-1 complex-valued neural network in their regions, respectively.

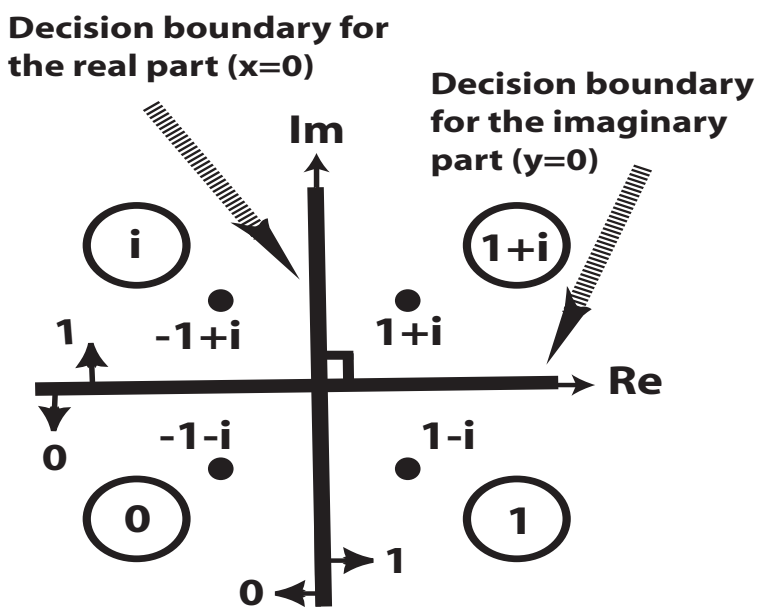

Figure 4: An example of the decision boundary of the 1-12-1 Complex-BP network learned with the learning pattern 1. The meanings of the numerals in Fig. 4 are as follows. **1**: Real part **OFF**(0.0-0.5), Imaginary part **OFF**, **2**: Real part **ON**(0.5-1.0), Imaginary part **OFF**, **3**: Real part **OFF**, Imaginary part **ON**, and **4**: Real part **ON**, Imaginary part **ON**. The decision boundary for the real part (i.e., the boundary that the region "**1+3**" and the region "**2+4**" form) and that for imaginary part (i.e., the boundary that the region "**1+2**" and the region "**3+4**" form) intersect orthogonally.

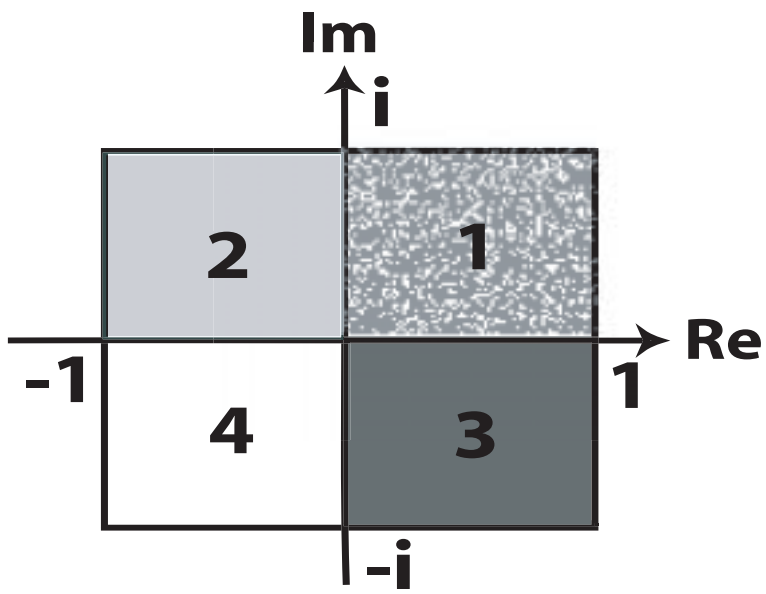

Figure 5: An example of the decision boundary of the 2-14-2 Real-BP network learned with the learning pattern 1. The numbers **1-4** have the same meanings as those of Fig. 4. The decision boundary for the real part (i.e., the boundary that the region "**1+3**" and the region "**2+4**" form) and that for imaginary part (i.e., the boundary that the region "**1+2**" and the region "**3+4**" form) do not intersect orthogonally.