Published in *International Journal of Neural Systems*, Vol.18, No.2, pp.123-134, 2008.

The Uniqueness Theorem for Complex-Valued Neural Networks with Threshold Parameters and the Redundancy of the Parameters

Tohru Nitta

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

This paper will prove the uniqueness theorem for 3-layered complex-valued neural networks where the threshold parameters of the hidden neurons can take non-zeros. That is, if a 3-layered complex-valued neural network is irreducible, the 3-layered complex-valued neural network that approximates a given complex-valued function is uniquely determined up to a finite group on the transformations of the learnable parameters of the complex-valued neural network.

1. Introduction

Complex-valued neural networks whose weights, threshold values, and input and output signals are all complex numbers, are useful from the viewpoint of applications. Actually, Watanabe et al. [1] applied the complex-valued back-propagation learning algorithm called *Complex-BP* [2,3,4] in the computer vision field. They successfully used the 2D motion learning ability (an ability to transform geometric figures) of the Complex-BP network to complement the *optical flow* which was a 2D velocity vector field derived from a set of images. Also, Miura et al. [5] applied the Complex-BP to the generation of *fractal images*. They used the 2D motion learning ability of the Complex-BP network to approximate iterated function systems (IFS). And, recently the problem of source separation has been extended to the complex domain [6,7,8], due to the need of frequency domain signal processing. The application fields of complex-valued neural networks have grown more and more.

It is important to clarify the properties of complex-valued neural networks in order to accelerate its practical applications. As for the inherent properties of complex-valued neural networks, several ones have already been shown in the literatures. [4,9] It should be noted that so-called component-wise complex-valued activation functions are used in the complex-valued neural networks described above. Although the Cauchy-Riemann equations do not hold, the component-wise activation function model has several distinct properties, compared with the real-valued neural network, and is a universal approximator. In recent years, non-component-wise complex-valued activation function models have been studied [10], which may have attractive properties that the component-wise activation function model does not have, although they are universal approximators conditionally.

The singularity is important because it affects the plateaus during learning [11], which causes a *standstill in learing*. This paper focuses on the singularity of complex-valued neural networks.

Sussmann investigated the class of 3-layered real-valued neural networks that can approximate a given real-valued function. He showed that the 3-layered real-valued neural network that can approximate the given real-valued function is uniquely determined up to a certain finite group (uniqueness theorem). [12] The extension of Sussmann's results to complex numbers is important for the reasons described below.

Fukumizu and Amari demonstrated the mechanism by which local minima of a real-valued neural network occur, based on Sussmann's results. [13] Hagiwara et al. showed theoretically that the AIC cannot be introduced into real-valued neural networks due to the nonuniqueness of the weights, based on Sussmann's results. [14] Furthermore, it follows from Sussmann's results that the Fisher information matrix for the parameters of the real-valued neural network is not positive-definite. In other words, the asymptotic regularity of maximum-likelihood estimation (i.e., asymptotically unbiased and minimum variance) of the parameters of the real-valued neural network is not guaranteed. Thus, the asymptotic behavior of maximum-likelihood estimation is unknown. In this context, Fukumizu investigated the asymptotic behavior of the maximum-likelihood estimation of unidentifiable models. [15] Watanabe rigorously derived an asymptotic representation for the stochastic complexity of an unidentifiable model and established a computational algorithm for its behavior. [16,17]

As seen above, Sussmann's results provide the basis for investigating the properties of real-valued neural networks, particularly the statistical properties. Similarly, it seems to be important that the properties of complex-valued neural networks corresponding to Sussmann's results should be revealed in order to have a clear knowledge of their characteristics.

As a matter of fact, we have already shown the singularity of complex-valued neural networks with hidden neurons whose thresholds are all zero, proving the uniqueness theorem in refs. 18 and 19. The assumption that thresholds of the hidden neurons are all zero is not, however, practical. Actually, the threshold parameters are not always zero in real applications. [1,5] So, the analysis for such a case is needed.

This paper proves the uniqueness theorem for complex-valued neural networks with hidden neurons whose thresholds are not always zero. That is, if a 3-layered complex-valued neural network is irreducible, the 3-layered complex-valued neural network that approximates a given complex-valued function is uniquely determined up to a certain finite group. The uniqueness theorem is basically proved by extending Sussmann's technique to complex numbers. Note that in the analysis of this paper, the hidden neurons are restricted to a certain family of hidden neurons, whereas such a restriction was not needed to prove the uniqueness theorem for complex-valued neural networks with hidden neurons whose thresholds are all zero. Such a restriction is caused by the technique for the proof which Sussmann used in ref. 12.

Section 2 describes the complex-valued neural network which is the object of analysis. Section 3 is devoted to derivation of the uniqueness theorem for complex-valued neural networks, which is followed by our conclusion in Section 4.

2. The Complex-Valued Neural Network

This section describes the 3-layered complex-valued neural network used in the analysis. 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 complex-valued neuron n. To obtain the (complex-valued) output signal, convert the net input U_n into its real and imaginary parts as follows:

$$U_n = x + iy = z, \tag{2}$$

where i denotes $\sqrt{-1}$. The (complex-valued) output signal is defined to be

$$\sigma(z) = \tanh(x) + i\tanh(y) \tag{3}$$

where $\tanh(u) \stackrel{\text{def}}{=} (\exp(u) - \exp(-u))/(\exp(u) + \exp(-u)), u \in \boldsymbol{R}$ ($\boldsymbol{R}$ denotes the set of real numbers) is called *hyperbolic tangent*. It is obvious that $-1 < Re[\sigma],\ Im[\sigma] < 1$. Note also that $\sigma(z)$ is not holomorphic, because the Cauchy-Riemann equations do not hold: $\partial Re[\sigma(z)]/\partial x \neq i\partial Im[\sigma(z)]/\partial y$ for any $z = x + iy \in \boldsymbol{C}$ such that $x \neq y$ ($\boldsymbol{C}$ denotes the set of complex numbers).

It seems that the way of the analysis used in this paper can be applied to the models with the complex-valued activation functions other than eqn (3). Kobayashi has addressed the uniqueness theorem for the model with the complex-valued activation function proposed by Hirose [20] $\sigma(z) = \tanh(|z|)\cdot(z/|z|)$ and confirmed that the uniqueness theorem holds true provided that all the thresholds of the hidden neurons are zero so far [21].

The complex-valued neural network consists of such complex-valued neurons described above. The network used in the analysis will have 3 layers: $m-n-1$ network. We will use $w_{kj} = w_{kj}^r + iw_{kj}^i \in \boldsymbol{C}$ for the weight between the input neuron j and the hidden neuron k, $c_k = c_k^r + ic_k^i \in \boldsymbol{C}$ for the weight between the hidden neuron k and the output neuron, $w_{k0} = w_{k0}^r + iw_{k0}^i \in \boldsymbol{C}$ for the threshold of the hidden neuron k, and $c_0 = c_0^r + ic_0^i \in \boldsymbol{C}$ for the threshold of the output neuron. Let $y_k(\boldsymbol{z}), h(\boldsymbol{z})$ denote the output values of the hidden neuron k, and the output neuron for the input pattern $\boldsymbol{z} = {}^t[z_1, \cdots, z_m] \in \boldsymbol{C}^m$ where $z_j = z_j^r + iz_j^i \in \boldsymbol{C}(1 \leq j \leq m)$, respectively. Let also $\nu_k(\boldsymbol{z})$ and $\mu(\boldsymbol{z})$ denote the net inputs to the hidden neuron k and the output neuron for the input pattern $\boldsymbol{z} \in \boldsymbol{C}^m$, respectively. That is, $\nu_k(\boldsymbol{z}) = \sum_{j=1}^m w_{kj}z_j + w_{k0}$, $\mu(\boldsymbol{z}) = \sum_{k=1}^n c_k y_k(\boldsymbol{z}) + c_0$, $y_k(\boldsymbol{z}) = \sigma(\nu_k(\boldsymbol{z}))$, and $h(\boldsymbol{z}) = \sigma(\mu(\boldsymbol{z}))$.

A constraint is imposed on the hidden neurons of the above complex-valued neural network, which we now state.

Definition 1 *If any of the following eight conditions is valid, the two complex-valued linear affine functions $\varphi_s : \boldsymbol{C}^m \to \boldsymbol{C}^1$ and $\varphi_t : \boldsymbol{C}^m \to \boldsymbol{C}^1$ are called* nearly rotation-equivalent*:*

$$
\begin{aligned}
\forall \boldsymbol{z} \in \boldsymbol{C}^m; Re[\varphi_s(\boldsymbol{z})] &= Re[\varphi_t(\boldsymbol{z})] \\
\textit{and} \ \exists \boldsymbol{z} \in \boldsymbol{C}^m; Im[\varphi_s(\boldsymbol{z})] &\neq Im[\varphi_t(\boldsymbol{z})], \quad (4) \\
\forall \boldsymbol{z} \in \boldsymbol{C}^m; Re[\varphi_s(\boldsymbol{z})] &= -Re[\varphi_t(\boldsymbol{z})] \\
\textit{and} \ \exists \boldsymbol{z} \in \boldsymbol{C}^m; Im[\varphi_s(\boldsymbol{z})] &\neq -Im[\varphi_t(\boldsymbol{z})], \quad (5) \\
\forall \boldsymbol{z} \in \boldsymbol{C}^m; Re[\varphi_s(\boldsymbol{z})] &= -Im[\varphi_t(\boldsymbol{z})] \\
\textit{and} \ \exists \boldsymbol{z} \in \boldsymbol{C}^m; Im[\varphi_s(\boldsymbol{z})] &\neq Re[\varphi_t(\boldsymbol{z})], \quad (6) \\
\forall \boldsymbol{z} \in \boldsymbol{C}^m; Re[\varphi_s(\boldsymbol{z})] &= Im[\varphi_t(\boldsymbol{z})] \\
\textit{and} \ \exists \boldsymbol{z} \in \boldsymbol{C}^m; Im[\varphi_s(\boldsymbol{z})] &\neq -Re[\varphi_t(\boldsymbol{z})], \quad (7) \\
\forall \boldsymbol{z} \in \boldsymbol{C}^m; Im[\varphi_s(\boldsymbol{z})] &= Im[\varphi_t(\boldsymbol{z})] \\
\textit{and} \ \exists \boldsymbol{z} \in \boldsymbol{C}^m; Re[\varphi_s(\boldsymbol{z})] &\neq Re[\varphi_t(\boldsymbol{z})], \quad (8) \\
\forall \boldsymbol{z} \in \boldsymbol{C}^m; Im[\varphi_s(\boldsymbol{z})] &= -Im[\varphi_t(\boldsymbol{z})] \\
\textit{and} \ \exists \boldsymbol{z} \in \boldsymbol{C}^m; Re[\varphi_s(\boldsymbol{z})] &\neq -Re[\varphi_t(\boldsymbol{z})], \quad (9) \\
\forall \boldsymbol{z} \in \boldsymbol{C}^m; Im[\varphi_s(\boldsymbol{z})] &= Re[\varphi_t(\boldsymbol{z})] \\
\textit{and} \ \exists \boldsymbol{z} \in \boldsymbol{C}^m; Re[\varphi_s(\boldsymbol{z})] &\neq -Im[\varphi_t(\boldsymbol{z})], \quad (10) \\
\forall \boldsymbol{z} \in \boldsymbol{C}^m; Im[\varphi_s(\boldsymbol{z})] &= -Re[\varphi_t(\boldsymbol{z})] \\
\textit{and} \ \exists \boldsymbol{z} \in \boldsymbol{C}^m; Re[\varphi_s(\boldsymbol{z})] &\neq Im[\varphi_t(\boldsymbol{z})]. \quad (11)
\end{aligned}
$$

(Remark) The net input $\varphi_k(\boldsymbol{z}) = \sum_{l=1}^m w_{kl}z_l + w_{k0}$ to a complex-valued neuron k is a complex-valued linear affine function on $\boldsymbol{C}^m$. The *nearly rotation-equivalence* is a concept that a complex-valued linear affine function on $\boldsymbol{C}^m$ comes close to be identically equal to another one by a counterclockwise rotation of it by a $0, \pi/2, \pi$ or $3\pi/2$ radians about the origin (See Definition 3).

Let $N_{m,n}$ be a set of all $m-n-1$ complex-valued neural networks such that the net inputs to any two hidden neurons are not nearly rotation-equivalent. $N_{m,n}$ is the object of the analysis in this paper.

Proposition 1 *The set of all complex-valued neural networks such that the thresholds of all hidden neurons are all zero, is included in $N_{m,n}$.*

Proof. Take any complex-valued neural network such that the thresholds of all hidden neurons are all zero, and fix it. And, take any two hidden neurons j and k from the complex-valued neural network. Since the thresholds are all zero, $\varphi_j(\boldsymbol{z}) \equiv \lambda_j(\boldsymbol{z})$ and $\varphi_k(\boldsymbol{z}) \equiv \lambda_k(\boldsymbol{z})$. Here, we can easily find that (a) $Re[\varphi_j(\boldsymbol{z})] \equiv Re[\varphi_k(\boldsymbol{z})] \iff Im[\varphi_j(\boldsymbol{z})] \equiv Im[\varphi_k(\boldsymbol{z})]$, (b) $Re[\varphi_j(\boldsymbol{z})] \equiv -Re[\varphi_k(\boldsymbol{z})] \iff Im[\varphi_j(\boldsymbol{z})] \equiv -Im[\varphi_k(\boldsymbol{z})]$, and (c) $Re[\varphi_j(\boldsymbol{z})] \equiv -Im[\varphi_k(\boldsymbol{z})] \iff Im[\varphi_j(\boldsymbol{z})] \equiv Re[\varphi_k(\boldsymbol{z})]$. So, φ_j and φ_k are not nearly rotation-equivalent. Thus, the proof is complete. □.

Proposition 1 states that the 3-layered complex-valued neural network with the hidden neurons whose thresholds are all zero, is the object of the analysis of this paper.

Proposition 2 *For any $n_1, n_2 \geq 1$ and for any $N_1 \in N_{m,n_1}, N_2 \in N_{m,n_2}$, the net input to any hidden neuron of N_1 and the one to any hidden neuron of N_2 are not nearly rotation-equivalent.*

Proof. It is trivial from the definition of the set $N_{m,n}$. □.

Proposition 2 is used in the proof of Theorem 1 (Uniqueness Theorem).

3. The Uniqueness Theorem for the Complex-Valued Neural Network

This section derives the uniqueness theorem for the complex-valued neural network described in the previous section. First, several definitions and preliminary propositions are presented, some of which are the same as those described in refs. 18 and 19 because the paper should be self-contained.

Definition 2 (I-O equivalence) *For a fixed m, two 3-layered complex-valued neural networks $N_1 \in N_{m,n_1}$ and $N_2 \in N_{m,n_2}$ are called* I-O equivalent *if their corresponding complex-valued functions are same.*

Definition 3 (rotation-equivalence) *Two complex-valued linear affine functions $\varphi_s, \varphi_t : \boldsymbol{C}^m \to \boldsymbol{C}^1$ are called* rotation-equivalent *if one of the following conditions holds:*

$$\varphi_s(\boldsymbol{z}) \equiv \varphi_t(\boldsymbol{z}) \quad \left(= e^{i0} \cdot \varphi_t(\boldsymbol{z})\right), \tag{12}$$

$$\varphi_s(\boldsymbol{z}) \equiv -\varphi_t(\boldsymbol{z}) \quad \left(= e^{i\pi} \cdot \varphi_t(\boldsymbol{z})\right), \tag{13}$$

$$\varphi_s(\boldsymbol{z}) \equiv i\varphi_t(\boldsymbol{z}) \quad \left(= e^{i\left(\frac{\pi}{2}\right)} \cdot \varphi_t(\boldsymbol{z})\right), \tag{14}$$

$$\varphi_s(\boldsymbol{z}) \equiv -i\varphi_t(\boldsymbol{z}) \quad \left(= e^{i\left(\frac{3\pi}{2}\right)} \cdot \varphi_t(\boldsymbol{z})\right). \tag{15}$$

(Remark) In addition to the two conditions of *sign-equivalence* that Sussmann defined to prove the uniqueness theorem for real-valued neural networks in ref. 12, the two new conditions (eqns (14) and (15)) related to the rotation of complex numbers are added.

Definition 4 (reducibility) *A complex-valued neural network $N \in N_{m,n}$ is called* reducible *if one of the following three conditions holds:*

1. *One of the weights between the hidden layer and the output layer is zero:*
   $1 \le \exists j \le n;\ c_j = 0$.
2. *There exist two hidden neurons such that the net inputs to them are rotation-equivalent:*
   $1 \le \exists j_1, \exists j_2 \le n;\ \varphi_{j_1}$ *and* φ_{j_2} *are rotation-equivalent.*
3. *There exists a hidden neuron such that the net input to it is a constant:*
   $1 \le \exists j \le n;\ \varphi_j$ *is a constant.*

As regards reducibility, the following property exits, as the name suggests.

Proposition 3 *If a 3-layered complex-valued neural network is reducible, then it is I-O equivalent to another 3-layered complex-valued neural network with fewer hidden neurons.*

Proof. See the appendix for the proof. Although this proposition can be proved in the same manner as ref. 19, the proof is presented in the appendix because the paper should be self-contained. □.

Definition 5 (irreducibility) *A 3-layered complex-valued neural network which is not reducible is called* irreducible*.*

The following four transformations are required in order to represent the redundancy of the learning parameters in a complex-valued neural network.

Proposition 4 *The following four transformations do not affect the complex-valued function realized by the 3-layered complex-valued neural network.*

1. *Sign inversion of learning parameters θ_j^{-1}*
   The transformation θ_j^{-1} reverses all of the learning parameters related to hidden neuron j, that is, the signs of all weights $\{w_{jk}\}_{k=1}^m$ between the input neuron k and the hidden neuron j, and the threshold w_{j0} of the hidden neuron j, as well as the weight c_j between the hidden neuron j and the output neuron.

2. *Correction of argument of learning parameters θ_j^i*
   The transformation θ_j^i multiplies each of the weights $\{w_{jk}\}_{k=1}^m$ between the input neuron k and the hidden neuron j, and the threshold w_{j0} of the hidden neuron j by $-i \in \boldsymbol{C}$, and also multiplies the weight c_j between the hidden neuron j and the output neuron by $i \in \boldsymbol{C}$.

3. *Correction of argument of learning parameters θ_j^{-i}*
   The transformation θ_j^{-i} multiplies each of the weights $\{w_{jk}\}_{k=1}^m$ between the input neuron k and the hidden neuron j, and the threshold w_{j0} of the hidden neuron j by $i \in \boldsymbol{C}$, and also multiplies the weight c_j between the hidden neuron j and the output neuron by $-i \in \boldsymbol{C}$.

4. *Exchange of hidden neurons $\tau_{j_1 j_2}$*
   The transformation $\tau_{j_1 j_2}$ exchanges two hidden neurons j_1 and j_2.

The 3-layered complex-valued neural networks which are obtained by applying the above four transformations to the 3-layered complex-valued neural network $N \in N_{m,n}$ are denoted as $\theta_j^{-1}(N), \theta_j^i(N), \theta_j^{-i}(N)$, and $\tau_{j_1 j_2}(N)$, respectively.

Proof. See the appendix for the proof. Although this proposition can be proved in the same manner as refs. 18 and 19, the proof is presented in the appendix because the paper should be self-contained. □.

Proposition 5 *The two transformations* θ_j^i *and* $\tau_{j_1 j_2}$ *described in Proposition 4 form the finite group* $W_{m,n}$ *related to the transformation of the set* $N_{m,n}$*. In other words,* $W_{m,n}$ *is the finite group formed by the transformations consisting of combinations of a finite number of the above two transformations:* $W_{m,n} = \{\theta_1^{-1}, \theta_2^i, \cdots, \theta_n^{-i}, \theta_1^i \circ \tau_{12}, \theta_1^{-i} \circ \tau_{23}, \cdots\}$.

Proof. Trivial □.

(Remark) $(\theta_j^i)^2 = \theta_j^{-1}$ and $(\theta_j^i)^3 = \theta_j^{-i}$.

Definition 6 (equivalence) *Consider two 3-layered complex-valued neural networks* $N_1 \in N_{m,n}$ *and* $N_2 \in N_{m,n}$*. If there exists a transformation* $h \in W_{m,n}$ *such that* N_1 *and* $h(N_2)$ *are the same 3-layered complex-valued neural networks (i.e., all of the learning parameters are the same),* $N_1 \in N_{m,n}$ *and* $N_2 \in N_{m,n}$ *are called* equivalent.

All the preparations have been done. The following Lemma 1 is required in order to prove Theorem 1 which is the main result obtained in this paper.

Lemma 1 *Let* J *be a finite set. Let* $\{\varphi_j\}_{j \in J}$ *be a family of complex-valued linear affine functions defined on* $\boldsymbol{C}^m$*, satisfying the following three conditions:*

1. *No function in* $\{\varphi_j\}_{j \in J}$ *is a constant.*
2. *No two functions in* $\{\varphi_j\}_{j \in J}$ *are rotation-equivalent.*
3. *No two functions in* $\{\varphi_j\}_{j \in J}$ *are nearly rotation-equivalent.*

Then, the complex-valued functions $\sigma \circ \varphi_j (j \in J)$ *and the constant function* 1 *are linearly independent.*

(Remark) $(\sigma \circ \varphi_j)(\boldsymbol{z}) \stackrel{\text{def}}{=} \sigma(\varphi_j(\boldsymbol{z}))$ for any $\boldsymbol{z} \in \boldsymbol{C}^m$.

Proof. Assume that $\breve{a} + \sum_{j \in J} a_j (\sigma \circ \varphi_j) \equiv 0$ where $\breve{a}, a_j \in \boldsymbol{C}$ $(j \in J)$. We have to prove that $\breve{a}$ and a_j are equal to zero for any $j \in J$.

Write $\varphi_j(z) = \lambda_j(z) + \lambda_j^0$ where λ_j is a nonzero complex-valued linear function and λ_j^0 is a complex-valued constant. Our hypothesis guarantees that, if $j_1 \neq j_2$, then either (a) the functions $\lambda_{j_1}, \lambda_{j_2}$ are not rotation-equivalent, or (b) if $\lambda_{j_1} \equiv \tau \lambda_{j_2}$ for some $\tau = \pm 1, \pm i$ then $\lambda_{j_1}^0 \neq \tau \lambda_{j_2}^0$ because φ_{j_1} and φ_{j_2} are not rotation-equivalent.

Define an equivalence relation on J by calling two elements j_1, j_2 of J equivalent if the corresponding linear complex-valued functions $\lambda_{j_1}, \lambda_{j_2}$ are rotation-equivalent. Let ε be the set of equivalence classes.

Pick a j_E for each $E \in \varepsilon$. As E varies over the classes in ε, no two of the functionals λ_{j_E} are rotation-equivalent. So, for each pair E_1, E_2 of distinct members of ε, the set of points $z \in \boldsymbol{C}^m$ where $\lambda_{j_{E_1}}(z) \neq \tau \lambda_{j_{E_2}}(z)$ for any $\tau = \pm 1, \pm i$ is open and dense in $\boldsymbol{C}^m$. Also, for each E, the set of $z \in \boldsymbol{C}^m$ such that $\lambda_{j_E}(z) \neq 0$ is open and dense in $\boldsymbol{C}^m$ because λ_{j_E} is not $\equiv 0$. So, we can pick an z such that $\lambda_{j_E}(z) \neq 0$ for all $E \in \varepsilon$, and $\lambda_{j_{E_1}}(z) \neq \tau \lambda_{j_{E_2}}(z)$ for any $\tau = \pm 1, \pm i$ and all pairs E_1, E_2 of distinct members of ε.

Here, let $\tilde{\lambda}_j(z) = \tau \lambda_j(z)$ for any $j \in J$ and $z \in \boldsymbol{C}^m$, $\tau = \pm 1, \pm i$ being chosen so that $Re[\tilde{\lambda}_j(z)] > 0$ and $Im[\tilde{\lambda}_j(z)] > 0$. Let $\tilde{\lambda}_j^0 = \tau \lambda_j^0$, $\tilde{\varphi}_j(z) = \tau \varphi_j(z)$ and $\tilde{a}_j = \tau^{-1} a_j$ with the same τ. Then,

$$\tilde{a} + \sum_{j \in J} \tilde{a}_j (\sigma \circ \tilde{\varphi}_j) \equiv 0. \tag{16}$$

Our proof will be complete if we show that $\tilde{a}$ and $\tilde{a}_j$ are equal to zero for any $j \in J$. Moreover, if $\tilde{a}_j$ is equal to zero for any $j \in J$, then eqn (16) shows that $\tilde{a} = 0$. So it will be enough to show that $\tilde{a}_j$ is equal to zero for any $j \in J$. Here, if $E \in \varepsilon$, all the $\tilde{\lambda}_j (j \in E)$ are equal to one and the same function, which we will call $\tilde{\lambda}_E$.

For any $j \in E$ and $z \in \boldsymbol{C}^m$,

$$\begin{aligned} \sigma(\tilde{\varphi}_j(z)) &= \frac{\xi_j^r \exp[2Re[\tilde{\lambda}_E(z)] - 1}{\xi_j^r \exp[2Re[\tilde{\lambda}_E(z)] + 1} \\ &\quad + i \frac{\xi_j^i \exp[2Im[\tilde{\lambda}_E(z)] - 1}{\xi_j^i \exp[2Im[\tilde{\lambda}_E(z)] + 1}, \end{aligned} \tag{17}$$

where $\xi_j^r = \exp[2Re[\tilde{\lambda}_j^0]]$ and $\xi_j^i = \exp[2Im[\tilde{\lambda}_j^0]]$. Since $Re[\tilde{\lambda}_E(z)] > 0$, $Im[\tilde{\lambda}_E(z)] > 0$, $\xi_j^r > 0$ and $\xi_j^i > 0$, we obtain $\lim_{t \to \infty} \sigma(\tilde{\varphi}_j(tz)) = 1 + i$. Thus, from eqn (16),

$$\tilde{a} + \sum_{j \in J} \tilde{a}_j (1 + i) \equiv 0. \tag{18}$$

If we subtract eqn (18) from eqn (16), we get

$$\sum_{j=1} \tilde{a}_j [\sigma \circ \tilde{\varphi}_j - (1 + i)] \equiv 0. \tag{19}$$

For any $z \in \boldsymbol{C}^m$,

$$(\sigma \circ \tilde{\varphi}_j)(z) - (1 + i) = -2\psi_j(z). \tag{20}$$

So, we have

$$\sum_{j\in J} \tilde{a}_j \psi_j \equiv 0, \qquad (21)$$

where $\psi_j = \psi_j^r + i\psi_j^i, \psi_j^r(z) = \frac{1}{\xi_j^r e^{2Re[\tilde{\lambda}_j(z)]}+1}$ and $\psi_j^i(z) = \frac{1}{\xi_j^i e^{2Im[\tilde{\lambda}_j(z)]}+1}$.

Now order the classes $E \in \varepsilon$ in a finite sequence $(E_1, E_2, \cdots, E_r)$, chosen so that $|\tilde{\lambda}_{E_1}(z)| < |\tilde{\lambda}_{E_2}(z)| < \cdots < |\tilde{\lambda}_{E_r}(z)|$. Let $v_k = \tilde{\lambda}_{E_k}(z), v_k^r = Re[\tilde{\lambda}_{E_k}(z)]$ and $v_k^i = Im[\tilde{\lambda}_{E_k}(z)]$. We then have

$$\psi_j(-tz) = \frac{1}{\xi_j^r e^{-2tv_k^r}+1} + i\frac{1}{\xi_j^i e^{-2tv_k^i}+1}, \qquad (22)$$

if $j \in E_k$. For each j, let $k(j)$ be the k such that $j \in E_k$. For $t > 0$ and sufficiently large, we have $0 < \xi_j^r e^{-2tv_k^r} < 1$ and $0 < \xi_j^i e^{-2tv_k^i} < 1$. So we can expand eqn (22) in a convergent power series:

$$\psi_j(-tz) = \sum_{s=0}^{\infty}(-\xi_j^r)^s e^{-2tsv_{k(j)}^r} + i\sum_{s=0}^{\infty}(-\xi_j^i)^s e^{-2tsv_{k(j)}^i}. \qquad (23)$$

If we multiply eqn (23) by $\tilde{a}_j$ and sum over j, we get

$$\begin{aligned} 0 &= \sum_{j\in J}\tilde{a}_j\psi_j(-tz) \qquad \text{(from eqn (21))} \\ &= \sum_{s=0}^{\infty}\sum_{k=1}^{r} e^{-2tsv_k^r}\sum_{j\in E_k}(-\xi_j^r)^s\tilde{a}_j \\ &\quad +i\sum_{s=0}^{\infty}\sum_{k=1}^{r} e^{-2tsv_k^i}\sum_{j\in E_k}(-\xi_j^i)^s\tilde{a}_j. \qquad (24) \end{aligned}$$

We rewrite eqn (24) as

$$\begin{aligned} 0 &= \sum_{\substack{v^r\in\boldsymbol{R} \\ v^r\geq 0}} e^{-2tv^r}\sum_{\substack{s\in\{0,1,\cdots\} \\ k\in\{1,2,\cdots,r\} \\ v^r=sv_k^r}}\sum_{j\in E_k}(-\xi_j^r)^s\tilde{a}_j \\ &\quad +i\sum_{\substack{v^i\in\boldsymbol{R} \\ v^i\geq 0}} e^{-2tv^i}\sum_{\substack{s\in\{0,1,\cdots\} \\ k\in\{1,2,\cdots,r\} \\ v^i=sv_k^i}}\sum_{j\in E_k}(-\xi_j^i)^s\tilde{a}_j. \end{aligned} \qquad (25)$$

The indexes of summation v^r and v^i are non-negative real numbers, but in fact the only v^r's and v^i's occurring in the summation are those that can be expressed as integral multiples of some v_k^r and v_k^i. So these v^r's and v^i's form discrete subsets $\triangle^r$ and $\triangle^i$ of the half-line $[0, \infty]$, respectively. If we order the elements of $\triangle^r \cup \triangle^i$ as a sequence $v^{\beta^{(1)}}, v^{\beta^{(2)}}, \cdots$, such that $0 \leq v^{\beta^{(1)}} \leq v^{\beta^{(2)}} \leq \cdots$ where β denotes r or i, then it follows easily from eqn (25) that for any $l = 1, 2, \cdots$,

$$\kappa_\beta^l \;=\; 0 \quad (\beta \text{ denotes } r \text{ or } i) \qquad (26)$$

where

$$\kappa_r^l = \sum_{\substack{s\in\{0,1,\cdots\} \\ k\in\{1,2,\cdots,r\} \\ sv_k^r=v^{r^{(l)}}}}\sum_{j\in E_k}(-\xi_j^r)^s\tilde{a}_j \quad (l = 1, 2, \cdots), \qquad (27)$$

$$\kappa_i^l = \sum_{\substack{s\in\{0,1,\cdots\} \\ k\in\{1,2,\cdots,r\} \\ sv_k^i=v^{i^{(l)}}}}\sum_{j\in E_k}(-\xi_j^i)^s\tilde{a}_j \quad (l = 1, 2, \cdots) \qquad (28)$$

(This can be proved by induction on $l = 1, 2, \cdots$).

It can be easily proved that $F \overset{\text{def}}{=} \{w \in \boldsymbol{C}^m \mid \tilde{\lambda}_{E_i}(w) \neq 0$ for any $i \in \{1, 2, \cdots, r\}$, $v_i^r(w) \neq v_j^r(w)$ for any $i, j \in \{1, 2, \cdots, r\}$ such that $i \neq j$, $v_i^i(w) \neq v_j^i(w)$ for any $i, j \in \{1, 2, \cdots, r\}$ such that $i \neq j$, $v_i^r(w) \neq v_j^i(w)$ for any $i, j \in \{1, 2, \cdots, r\}$ such that $i \neq j$, $v_i^r(w) \neq v_i^i(w)$ for any $i \in \{1, 2, \cdots, r\}$ $\}$ is not empty. Then, choose a $w \in F$, and let $v_k^r = v_k^r(w), v_k^i = v_k^i(w)$ $(1 \leq k \leq r)$. Fix a $k \in \{1, 2, \cdots, r\}$, and let $\alpha = |E_k|$. Then, Choose integers $s^r > 0, h^r > 0$, such that $s^r v_k^r, (s^r + h^r)v_k^r, (s^r + 2h^r)v_k^r, \cdots, (s^r + (\alpha-1)h^r)v_k^r$ are not integral multiples of $v_{k'}^r (k < k' \leq r)$ and $v_{k'}^i (k \leq k' \leq r)$ (See the appendix for the proof).

Assume that we have already proved that $\tilde{a}_j$ is equal to zero for any $j \in E_{k'}, k' < k$. Let δ^r be an integer such that $0 \leq \delta^r < \alpha$. Then, the following equations hold true:

$$\sum_{j\in E_k}(-\xi_j^r)^{s^r+\delta^r h^r}\tilde{a}_j = 0 \quad (\delta^r = 0, 1, \cdots, \alpha-1). \qquad (29)$$

Eqn (29) can be proved as follows. First, there exists some l such that $(s^r + \delta^r h^r)v_k^r = v^{r^{(l)}}$. For the above l,

$$\kappa_r^l = \sum_{\substack{s'\in\{0,1,\cdots\} \\ k'\in\{1,2,\cdots,r\} \\ s'v_{k'}^r=v^{r^{(l)}}}}\sum_{j\in E_{k'}}(-\xi_j^r)^{s'}\tilde{a}_j \quad (l = 1, 2, \cdots). \qquad (30)$$

This sum contains no contribution coming from values of k' such that $k' < k$, because we are assuming that all the corresponding $\tilde{a}_j$ vanish. A k' such that $k' > k$ can only contribute to the sum if $s'v_{k'}^r = v^{r^{(l)}}$ for some integer s', and this can only happen if $s'v_{k'}^r = (s^r + \delta^r h^r)v_k^r$. Since this is impossible by our choice of s^r and h^r, we conclude that the sum only contains contributions coming from $k' = k$. In this case, the only possible value of s' is $s + \delta h$. So,

$$\begin{aligned} 0 &= \kappa_r^l \qquad\qquad \text{(from eqn(26))} \\ &= \sum_{j \in E_k} (-\xi_j^r)^{s^r + \delta^r h^r} \tilde{a}_j. \end{aligned} \tag{31}$$

Thus, eqn (29) is proved.

Here, eqn (29) can be rewritten as follows:

$$\begin{aligned} 0 &= \sum_{j \in E_k} (-\xi_j^r)^{s^r + \delta^r h^r} \tilde{a}_j \\ &= \sum_{j \in E_k} g_j(\delta^r) b_j \quad (\delta^r = 0, 1, \cdots, \alpha - 1), \end{aligned} \tag{32}$$

where

$$g_j(\delta^r) = (-\xi_j^r)^{\delta^r h^r} \quad (j \in E_k), \tag{33}$$

$$b_j = \tilde{a}_j (-\xi_j^r)^{s^r} \quad (j \in E_k). \tag{34}$$

Furthermore, eqn (32) can be rewritten as follows:

$$\begin{bmatrix} 1 & \cdots & 1 \\ g_1(1) & \cdots & g_\alpha(1) \\ (g_1(1))^2 & \cdots & (g_\alpha(1))^2 \\ \vdots & & \vdots \\ (g_1(1))^{\alpha-1} & \cdots & (g_\alpha(1))^{\alpha-1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_\alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{35}$$

because $g_j(0) = 1$, $g_j(1) = (-\xi_j^r)^{h^r}$, $g_j(2) = (g_j(1))^2, \cdots, g_j(\alpha-1) = (g_j(1))^{\alpha-1}$. The matrix (we let it denote X) in eqn (35) is a Vandermonde matrix whose determinant is given by $\Pi_{l>m}(g_l(1) - g_m(1))$.

Here, for any $k \in \{1, 2, \cdots, r\}$, for any $l, m \in E_k$ such that $l \neq m$,

$$g_l(1) \neq g_m(1). \tag{36}$$

Eqn (36) can be proved as follows. Fix $k \in \{1, 2, \cdots, r\}$ arbitrarily. Also, fix $l, m \in E_k$ such that $l \neq m$ arbitrarily. Since $g_l(1) = (-\xi_l^r)^{h^r}$ and $\xi_l^r = e^{2Re[\tilde{\lambda}_l^0]}$, the proof of eqn (36) is complete if $Re[\tilde{\lambda}_l^0] \neq Re[\tilde{\lambda}_m^0]$ is shown. Since $l, m \in E_k$, there exists some $\tau = \pm 1, \pm i$ such that $\lambda_l \equiv \tau \lambda_m$. So, $\tilde{\lambda}_l \equiv \tilde{\lambda}_m$. Here, assume that $Re[\tilde{\lambda}_l^0] = Re[\tilde{\lambda}_m^0]$. Then, from the assumption of Lemma 1, φ_l and φ_m are not nearly rotation-equivalent. Then, the following equation holds:

$$Im[\tilde{\lambda}_l^0] = Im[\tilde{\lambda}_m^0] \tag{37}$$

(See the appendix for the proof). Thus, $\tilde{\varphi}_l \equiv \tilde{\varphi}_m$. So, φ_l and φ_m are rotation-equivalent, which contradicts the assumption of Lemma 1. This means that $Re[\tilde{\lambda}_l^0] \neq Re[\tilde{\lambda}_m^0]$. Thus, eqn (36) is proved.

From eqn (36), $\det X = \Pi_{l>m}(g_l(1) - g_m(1)) \neq 0$. So, we conclude that all the b_j vanish. However, this implies that all the $\tilde{a_j}$ vanish as well because $b_j = \tilde{a}_j(-\xi_j^r)^{s^r}$ and $\xi_j^r = e^{2Re[\tilde{\lambda}_j^0]} > 0$. □.

The following Theorem 1 is the main result obtained in this paper.

Theorem 1 (Uniqueness Theorem) *Let N_1, N_2 be irreducible I-O equivalent 3-layered complex-valued neural networks in N_{m,n_1}, N_{m,n_2}, respectively. Then (i) $n_1 = n_2$ and (ii) N_1 and N_2 are equivalent.*

Theorem 1 states that the redundancy of the learning parameters of an irreducible complex-valued neural network to approximate a given complex-valued function is determined up to a finite group.

Proof. Consider two complex-valued neural networks $N_1 \in N_{m,n_1}$ and $N_2 \in N_{m,n_2}$ which are irreducible and I-O equivalent. Suppose that $n_1 \leq n_2$ (generality is not lost by this assumption). Since N_1 and N_2 are I-O equivalent, we have $h^1(\boldsymbol{z}) = h^2(\boldsymbol{z})$ for any input pattern $\boldsymbol{z} \in \boldsymbol{C}^m$. In explicit form, the following relation is obtained:

$$(c_0^1 - c_0^2) + \sum_{j=1}^{n_1} c_j^1 \sigma(\nu_j^1(\boldsymbol{z})) - \sum_{j=1}^{n_2} c_j^2 \sigma(\nu_j^2(\boldsymbol{z})) = 0. \tag{38}$$

The variables with superscript 1 are concerned with the complex-valued neural network N_1, and the variables with superscript 2 are concerned with the complex-valued neural network N_2.

Let $a_0 = c_0^1 - c_0^2, J = \{1, 2, \cdots, n_1 + n_2\}, a_j = c_j^1 (1 \leq j \leq n_1), a_j = -c_{j-n_1}^2 (n_1 + 1 \leq j \leq n_1 + n_2), \varphi_j = \nu_j^1 (1 \leq j \leq n_1)$ and $\varphi_j = \nu_{j-n_1}^2 (n_1 + 1 \leq j \leq n_1 + n_2)$. Then, eqn (38) is written as

$$a_0 + \sum_{j \in J} a_j \sigma(\varphi_j(\boldsymbol{z})) = 0. \tag{39}$$

Since N_1 and N_2 are irreducible, it cannot occur that all of $a_1, \cdots, a_{n_1+n_2} \in \boldsymbol{C}$ are 0. Consequently, by Lemma 1, at least one of the following is valid: (a) there exists $j \in J$ such that φ_j is a constant function, (b) there exist $1 \leq j_1,\ j_2 \leq n_1 + n_2$ such that φ_{j_1} and φ_{j_2} are rotation-equivalent, (c) there exist $1 \leq j_1,\ j_2 \leq n_1 + n_2$ such that φ_{j_1} and φ_{j_2} are nearly rotation-equivalent.

Since N_1 and N_2 are irreducible, however, (a) cannot be the case. And, from the proposition 2, (c) cannot be the case. Thus, (b) must be true.

It also follows from the irreducibility of N_1 and N_2 that φ_{j_1} and φ_{j_2} are not of the same network. Thus, it must be the case that $1 \le j_1 \le n_1$ and $n_1+1 \le j_2 \le n_1+n_2$. Eqn (39) is therefore written as

$$(a_{j_1} + \rho a_{j_2})\sigma(\varphi_{j_1}(\boldsymbol{z})) + \left[a_0 + \sum_{\substack{j\in J \\ j\neq j_1,j_2}} a_j\sigma(\varphi_j(\boldsymbol{z})) \right] = 0 \quad (40)$$

where $\rho = \pm 1, \pm i$. Since N_1 and N_2 are irreducible, any $\varphi_j (j \neq j_1, j_2)$ appearing in eqn (40) is not rotation-equivalent to φ_{j_1}.

Assume that $a_{j_1} + \rho a_{j_2} \neq 0$. Then $\sigma \circ \varphi_{j_1}$, 1 (constant function) and $\{\sigma \circ \varphi_j\}_{j\in J, j\neq j_1,j_2}$ are linearly dependent. Then, by Lemma 1, at least one of the following is valid: (a) there exists $k \in J$ such that $k \neq j_2$ and φ_k is a constant function, (b) there exist $1 \le k, l \le n_1+n_2$ such that $k, l \neq j_2$, and φ_k and φ_l are rotation-equivalent, (c) there exist $1 \le k, l \le n_1+n_2$ such that $k, l \neq j_2$, and φ_k and φ_l are nearly rotation-equivalent. Since N_1 and N_2 are irreducible, however, (a) cannot be the case. And, from the proposition 2, (c) cannot be the case. Thus, (b) must be true. The two can be integrated as follows:

$$a_k\sigma(\varphi_k(\boldsymbol{z})) + a_l\sigma(\varphi_l(\boldsymbol{z})) = (a_k + \rho' a_l)\sigma(\varphi_k(\boldsymbol{z})) \quad (41)$$

where $\rho' = \pm 1, \pm i$. Here, φ_k is not rotation-equivalent to any $\varphi_j (j \in J,\ j \neq j_1, j_2, k, l)$ (due to the irreducibility of N_1 and N_2). Thus, still $\sigma \circ \varphi_{j_1}$, 1 and $\{\sigma \circ \varphi_j\}_{j\in J,\ j\neq j_1,j_2,l}$ are linearly dependent. Then, by Lemma 1, at least one of the following is valid: (a) there exists $s \in J$ such that $s \neq j_2, l$ and φ_s is a constant function, (b) there exist $1 \le s, t \le n_1+n_2$ such that $s, t \neq j_2, l$, and φ_s and φ_t are rotation-equivalent, (c) there exist $1 \le s, t \le n_1+n_2$ such that $s, t \neq j_2, l$, and φ_s and φ_t are nearly rotation-equivalent. Since N_1 and N_2 are irreducible, (a) cannot be the case. And, from the proposition 2, (c) cannot be the case. Thus, (b) must be true. By iterating the same procedure, a family of functions $\{\varphi_j\}_{j\in K}$ is constructed from $\{\varphi_j\}_{j\in J, j\neq j_2}$, in which no two φ_j and $\varphi_{j'}$ are rotation-equivalent. Then, φ_{j_1} and some $\varphi_j, j \in K$ must be rotation-equivalent. This contradicts the statement below eqn (40). Thus, we have $a_{j_1} + \rho a_{j_1} = 0$.

Eqn (40) is therefore written as

$$a_0 + \sum_{\substack{j\in J \\ j\neq j_1,j_2}} a_j\sigma(\varphi_j(\boldsymbol{z})) = 0. \quad (42)$$

Applying the same procedure as above to eqn (42), it follows that there exist $1 \le j_3 \le n_1$ such that $j_3 \neq j_1$ and $n_1+1 \le j_4 \le n_1+n_2$ such that $j_4 \neq j_2$ such that φ_{j_3} and φ_{j_4} are rotation-equivalent, and the following relation is valid:

$$(a_{j_3} + \rho' a_{j_4})\sigma(\varphi_{j_3}(\boldsymbol{z})) + \left[a_0 + \sum_{\substack{j\in J \\ j\neq j_1,j_2,j_3,j_4}} a_j\sigma(\varphi_j(\boldsymbol{z})) \right] = 0, \quad (43)$$

$$a_{j_3} + \rho' a_{j_4} = 0, \quad (44)$$

$$a_0 + \sum_{\substack{j\in J \\ j\neq j_1,j_2,j_3,j_4}} a_j\sigma(\varphi_j(\boldsymbol{z})) = 0 \quad (45)$$

where $\rho' = \pm 1, \pm i$. Since N_1 and N_2 are irreducible, any $\varphi_j (j \neq j_1, j_2, j_3, j_4)$ appearing in eqn (45) is not rotation-equivalent to φ_{j_3}.

Continuing the same procedure, we see that there must exist $\varphi_{j'} (n_1+1 \le j' \le n_1+n_2)$ for any $1 \le j \le n_1$ which is rotation-equivalent to φ_j.

Assuming that $n_1 < n_2$, we finally obtain

$$a_0 + \sum_{\substack{j\in J \\ j\ge n_1+1}} a_j\sigma(\varphi_j(\boldsymbol{z})) = 0. \quad (46)$$

Since none of $a_j (j \in J, j \ge n_1+1)$ are zero, one of the following conditions is valid, as can be seen from Lemma 1: (a) there exists some $n_1+1 \le j \le n_1+n_2$ such that φ_j is a constant function, (b) there exist $n_1+1 \le j,\ j' \le n_1+n_2$ such that φ_j and $\varphi_{j'}$ are rotation-equivalent, (c) there exist $n_1+1 \le j,\ j' \le n_1+n_2$ such that φ_j and $\varphi_{j'}$ are nearly rotation-equivalent. Here, (a) and (b) contradict the irreducibility of N_2. (c) cannot be the case becasue of the proposition 2. Consequently, it must be the case that $n_1 = n_2$.

Summarizing, the following relations are obtained:

$$a_0 + \sum_{l=1}^{n_1} (a_{j_l}^{(1)} + \rho_l a_{j_l}^{(2)})\sigma(\varphi_{j_l}^{(1)}(\boldsymbol{z})) = 0, \quad (47)$$

$$a_0 = 0, \quad (48)$$

$$a_{j_l}^{(1)} + \rho_l a_{j_l}^{(2)} = 0 \qquad (1 \le l \le n_1) \quad (49)$$

where $\rho_l = \pm 1, \pm i$ $(1 \le l \le n_1)$. The coefficient and function with superscript "(1)" are concerned with N_1, and the coefficient with superscript "(2)" are concerned with N_2. No two of $\{\varphi_{j_l}^{(1)}\}_{l=1}^{n_1}$ appearing eqn (47) are rotation-equivalent (this is obvious from the construction).

First, it follows from $a_0 = c_0^1 - c_0^2 = 0$ (eqn (48)) that $c_0^1 = c_0^2$ (the thresholds of the output neurons in N_1 and N_2 are equal). It follows from eqn (49) that one of the following four relations is valid for any $1 \le l \le n_1$:

$$a_{j_l}^{(1)} + a_{j_l}^{(2)} = 0 \text{ and } \varphi_{j_l}^{(1)}(\boldsymbol{z}) = \varphi_{j_l}^{(2)}(\boldsymbol{z}), \quad (50)$$

$$a_{j_l}^{(1)} - a_{j_l}^{(2)} = 0 \text{ and } \varphi_{j_l}^{(1)}(\boldsymbol{z}) = -\varphi_{j_l}^{(2)}(\boldsymbol{z}), \quad (51)$$

$$a_{j_l}^{(1)} + ia_{j_l}^{(2)} = 0 \text{ and } \varphi_{j_l}^{(1)}(\boldsymbol{z}) = -i\varphi_{j_l}^{(2)}(\boldsymbol{z}), \quad (52)$$

$$a_{j_l}^{(1)} - ia_{j_l}^{(2)} = 0 \text{ and } \varphi_{j_l}^{(1)}(\boldsymbol{z}) = i\varphi_{j_l}^{(2)}(\boldsymbol{z}). \quad (53)$$

Each of these four cases is considered separately.

1. Case of eqn (50)
   It follows from $a_{j_l}^{(1)} + a_{j_l}^{(2)} = 0$ that $c_{j_l}^{(1)} = c_{j_l}^{(2)}$, that is, the weight between the hidden neuron j_l and the output neuron in N_1 and that in N_2 are equal. It follows from $\varphi_{j_l}^{(1)}(\boldsymbol{z}) = \varphi_{j_l}^{(2)}(\boldsymbol{z})$ that $\nu_{j_l}^1(\boldsymbol{z}) = \nu_{j_l}^2(\boldsymbol{z})$ for any $\boldsymbol{z} \in \boldsymbol{C}^m$, that is, $\sum_{k=1}^m w_{kj_l}^1 z_k + w_{j_l0}^1 = \sum_{k=1}^m w_{kj_l}^2 z_k + w_{j_l0}^2$. Thus, $w_{kj_l}^1 = w_{kj_l}^2 (1 \le k \le m)$ and $w_{j_l0}^1 = w_{j_l0}^2$. In other words, the weight between the hidden neuron j_l and the input neuron in N_1 is equal to that in N_2, and the threshold of the hidden neuron j_l in N_1 is equal to that in N_2.

2. Case of eqn (51)
   It follows from $a_{j_l}^{(1)} - a_{j_l}^{(2)} = 0$ that $c_{j_l}^{(1)} = -c_{j_l}^{(2)}$, that is, the weight between the hidden neuron j_l and the output neuron in N_1 is equal to that in N_2 with reversal of sign. It also follows from $\varphi_{j_l}^{(1)}(\boldsymbol{z}) = -\varphi_{j_l}^{(2)}(\boldsymbol{z})$ that $\nu_{j_l}^1(\boldsymbol{z}) = -\nu_{j_l}^2(\boldsymbol{z})$ for any $\boldsymbol{z} \in \boldsymbol{C}^m$, that is, $\sum_{k=1}^m w_{kj_l}^1 z_k + w_{j_l0}^1 = \sum_{k=1}^m (-w_{kj_l}^2) z_k - w_{j_l0}^2$. Consequently, $w_{kj_l}^1 = -w_{kj_l}^2 (1 \le k \le m)$ and $w_{j_l0}^1 = -w_{j_l0}^2$. In other words, the weight between the hidden neuron j_l and the input neuron in N_1 is equal to that in N_2 with reversal of sign, and the threshold of the hidden neuron j_l in N_1 is equal to that in N_2 with reversal of sign. Consequently, the result of applying the transformation $\theta_{j_l}^{-1}$ to N_2 is equal to N_1, that is, $\theta_{j_l}^{-1}(N_2) = N_1$.

3. Case of eqn (52)
   It follows from $a_{j_l}^{(1)} + ia_{j_l}^{(2)} = 0$ that $c_{j_l}^{(1)} = ic_{j_l}^{(2)}$, that is, the weight between the hidden neuron j_l and the output neuron in N_1 is equal to that in N_2 multiplied by i. It also follows from $\varphi_{j_l}^{(1)}(\boldsymbol{z}) = -i\varphi_{j_l}^{(2)}(\boldsymbol{z})$ that $\nu_{j_l}^1(\boldsymbol{z}) = -i\nu_{j_l}^2(\boldsymbol{z})$ for any $\boldsymbol{z} \in \boldsymbol{C}^m$, that is, $\sum_{k=1}^m w_{kj_l}^1 z_k + w_{j_l0}^1 = \sum_{k=1}^m (-iw_{kj_l}^2) z_k - iw_{j_l0}^2$. Consequently, $w_{kj_l}^1 = -iw_{kj_l}^2 (1 \le k \le m)$ and $w_{j_l0}^1 = -iw_{j_l0}^2$. In other words, the weight between the hidden neuron j_l and the input neuron in N_1 is equal to that in N_2 multiplied by $-i$, and the threshold of the hidden neuron j_l in N_1 is equal to that in N_2 multiplied by $-i$. In other words, $\theta_{j_l}^{i}(N_2) = N_1$.

4. Case of eqn (53)
   It follows from $a_{j_l}^{(1)} - ia_{j_l}^{(2)} = 0$ that $c_{j_l}^{(1)} = -ic_{j_l}^{(2)}$, that is, the weight between the hidden neuron j_l and the output neuron in N_1 is equal to that in N_2 multiplied by $-i$. It also follows from $\varphi_{j_l}^{(1)}(\boldsymbol{z}) = i\varphi_{j_l}^{(2)}(\boldsymbol{z})$ that $\nu_{j_l}^1(\boldsymbol{z}) = i\nu_{j_l}^2(\boldsymbol{z})$ for any $\boldsymbol{z} \in \boldsymbol{C}^m$, that is, $\sum_{k=1}^m w_{kj_l}^1 z_k + w_{j_l0}^1 = \sum_{k=1}^m (iw_{kj_l}^2) z_k + iw_{j_l0}^2$. Consequently, $w_{kj_l}^1 = iw_{kj_l}^2 (1 \le k \le m)$ and $w_{j_l0}^1 = iw_{j_l0}^2$. In other words, the weight between the hidden neuron j_l and the input neuron in N_1 is equal to that in N_2 multiplied by i, and the threshold of the hidden neuron j_l in N_1 is equal to that in N_2 multiplied by i. In other words, $\theta_{j_l}^{-i}(N_2) = N_1$.

Thus, N_1 and N_2 are equivalent. □.

Theorem 2 can be proved in the same manner as ref. 19. See the appendix for the proof.

Theorem 2 *The order of the finite group $W_{m,n}$ described in Proposition 5 is $2^{2n} \cdot n!$.*

Definition 7 (minimality) *If a three-layered complex-valued neural network with n hidden neurons is not I-O equivalent to any three-layered complex-valued neural network with $n-1$ or fewer hidden neurons, that three-layered complex-valued neural network is called* minimal.

Corollary 1 *The irreducible three-layered complex-valued neural network is minimal.*

The results obtained in this paper are useful from an engineering viewpoint. For example, when a large-scale complex-valued neural network is used to process large-scale image data, there is a possibility that the network is reducible. It is desirable for processing efficiency that the network employed be as small in size as possible. After learning, the minimal network can be obtained by checking the three conditions of Definition 4 one by one. Corollary 1 guarantees this.

4. Conclusions

This paper has proved the uniqueness theorem for the complex-valued neural network. That is, if a 3-layered complex-valued neural network is irreducible, the 3-layered complex-valued neural network that approximates a given complex-valued function is uniquely determined up to a certain finite group. The assumption for the analysis is more practical than those of the previous papers, [18,19] that is, this paper has dealt with the 3-layered complex-valued neural network with the hidden neurons whose thresholds are not always zero, whereas the previous papers [18,19] dealt with the one with the hidden neurons whose thresholds were all zero. We believe that the results can be the basis for investigating the properties of complex-valued neural networks, particularly the statistical properties.

Note that the constraint on *nearly rotation-equivalent* was imposed on the hidden neurons of the complex-valued neural network. As far as Sussmann's technique for proof is used, such a constraint inevitably occurs, which seems to be caused by the non-holomorphic activation function (eqn (3)) of the complex-valued neuron. The other technique for proof would be needed in order to remove the constraint on nearly rotation-equivalent.

Thimm and Moerland [22] proved that there existed a precise relationship between gain parameter (i.e., slope in the nonlinear activation function), learning rate, and initial weights for two backpropagation feedforward neural networks M and N with the same topology: the neural network M whose gain parameter is β, learning rate η, and initial weights $\boldsymbol{w}$ is equivalent to the neural network N whose gain parameter is 1, learning rate $\beta^2\eta$, and initial weights $\beta\boldsymbol{w}$. Furthermore, Mandic and Chambers extended the results obtained by Thimm and Moerland to a class of recurrent neural networks trained by the real-time recurrent learning algorithm [23,24]. We believe that the uniqueness theorem proved in this paper can be generalized using those results.

In the future, the asymptotic behavior of the maximum likelihood estimator of the learning parameters in the complex-valued neural network will be analyzed on the basis of the results presented in this paper.

Acknowledgements

We wish to thank the members of the Mathematical Neuroinformatics Group for discussing this study and the anonymous reviewers for valuable comments.

Appendix

Proof of Proposition 3:
Suppose that a three-layered complex-valued neural network is reducible. Then, one of the three conditions in Definition 4 is valid. If the first condition is valid, the output of hidden neuron j has no effect on the output of the output neuron. Consequently, the hidden neuron j can be eliminated. If the second condition is valid, let the weight between hidden neuron j_1 and the output neuron be c_{j_1}, and the weight between the hidden neuron j_2 and the output neuron be c_{j_2}. Then, the hidden neuron j_2 is eliminated and the weight between hidden neuron j_1 and the output neuron is modified from c_{j_1} to $c_{j_1}+\rho c_{j_2}$. Then the resulting complex-valued neural network is I-O equivalent to the original complex-valued neural network. Here, ρ is a value that depends on which of the four conditions (eqns (12) - (15)) for rotation-equivalence is satisfied; it takes one of the values $1, -1, i$ and $-i$. If the third condition is valid, the output of the hidden neuron j is always constant. Consequently, the value is added to the threshold of the output neuron and hidden neuron j is eliminated. Then the complex-valued neural network is I-O equivalent to the original complex-valued neural network.

Proof of Proposition 4:
It is noted that tanh is an odd function, that is, $\tanh(-u) = -\tanh(u)$. The following properties of complex numbers are also noted. For any complex numbers $w, z \in \boldsymbol{C}$, we have $Re[(-w)z] = -Re[wz]$, $Im[(-w)z] = -Im[wz]$, $Re[(-i)z] = Im[z]$, $Im[(-i)z] = -Re[z]$, $Re[iz] = -Im[z]$ and $Im[iz] = Re[z]$. Thus, it is easy to show that the output of the output neuron is not affected by any of the four transformations.

Proof of choosing integers $s^r > 0$ and $h^r > 0$:
Let B be the subset of $\{k+1, \cdots, r\}$ consisting of those k' such that the quotient $v^r_{k'}/v^r_k$ is rational.

Write this quotient as $p_{k'}/q_{k'}$, where $p_{k'}$ and $q_{k'}$ are relatively prime positive integers. Also, let C be the subset of $\{k, \cdots, r\}$ consisting of those k'' such that the quotient $v^i_{k''}/v^r_k$ is rational. Write this quotient as $p_{k''}/q_{k''}$, where $p_{k''}$ and $q_{k''}$ are relatively prime positive integers. Then, sv^r_k cannot be an integer multiple of $v^r_{k'}$ unless s is divisible by $p_{k'}$ where s is a positive integer (This can be proved as follows. Assume that there exists an integer n such that $sv^r_k = nv^r_{k'}$. Then, $s = nv^r_{k'}/v^r_k = np_{k'}/q_{k'}$. So, $n = sq_{k'}/p_{k'}$. Thus, s must be divisible by $p_{k'}$.) Similarly, sv^r_k cannot be an integer multiple of $v^i_{k''}$ unless s is divisible by $p_{k''}$ where s is a positive integer. So, if we pick $h^r = \Pi_{k' \in B, k'' \in C} p_{k'} p_{k''}$ and $s^r = 1 + h^r$, we see that $s^r v^r_k$ cannot be an integer multiple of $v^r_{k'}$ and $v^i_{k''}$ for any $k' \in B$ and $k'' \in C$. If δ is an integer, then $(s^r + \delta h^r)v^r_k$ cannot be an integer multiple of $v^r_{k'}$ and $v^i_{k''}$ for any $k' \in B$ and $k'' \in C$ because $s^r + \delta h^r$ can never be divisible by $p_{k'}$ and $p_{k''}$. So s^r and h^r are the desired integers. On the other hand, if $v^r_{k'}/v^r_k$ is irrational, $(s^r + \delta h^r)v^r_k$ cannot be an integer multiple of $v^r_{k'}$ (This can be proved as follows. Let $x = v^r_{k'}/v^r_k$. Then, $v^r_{k'} = xv^r_k$. So, $(s^r + \delta h^r)v^r_{k'} = (s^r + \delta h^r)xv^r_k$. Thus, $(s^r + \delta h^r)v^r_k = [(s^r + \delta h^r)/x]v^r_{k'}$ where $(s^r + \delta h^r)/x$ is not an integer). Similarly, if $v^i_{k''}/v^r_k$ is irrational, $(s^r + \delta h^r)v^r_k$ cannot be an integer multiple of $v^i_{k''}$.

Proof of eqn (37):
Since there exist some $\tau_l, \tau_m = \pm 1, \pm i$ such that $\widetilde{\lambda}_l = \tau_l \lambda_l$ and $\widetilde{\lambda}_m = \tau_m \lambda_m$, we will show $Im[\widetilde{\lambda}^0_l] = Im[\widetilde{\lambda}^0_m]$ for each case of $\tau_l, \tau_m = \pm 1, \pm i$. In case of $\tau_l = \tau_m = 1$, $\lambda_l \equiv \lambda_m$ (i.e., $Re[\lambda_l] \equiv Re[\lambda_m], Im[\lambda_l] \equiv Im[\lambda_m]$) and $Re[\lambda^0_l] = Re[\lambda^0_m]$. So, for any $\boldsymbol{z} \in \boldsymbol{C}^m$, $Re[\varphi_l(\boldsymbol{z})] = Re[\lambda_l(\boldsymbol{z})] + Re[\lambda^0_l] = Re[\lambda_m(\boldsymbol{z})] + Re[\lambda^0_m] = Re[\varphi_m(\boldsymbol{z})]$. Thus, from eqn (4), $Im[\varphi_l] \equiv Im[\varphi_m]$ holds true. Then, for any $\boldsymbol{z} \in \boldsymbol{C}^m$, $Im[\varphi_l(\boldsymbol{z})] = Im[\lambda_l(\boldsymbol{z})] + Im[\lambda^0_l] = Im[\lambda_m(\boldsymbol{z})] + Im[\lambda^0_l] = Im[\varphi_m(\boldsymbol{z})] + Im[\lambda^0_l] - Im[\lambda^0_m]$. Thus, $Im[\lambda^0_l] = Im[\lambda^0_m]$. Therefore, $Im[\widetilde{\lambda}^0_l] = Im[\lambda^0_l] = Im[\lambda^0_m] = Im[\widetilde{\lambda}^0_m]$. The other cases can be shown in the same manner.

Proof of Theorem 2:
$|W_{m,n}|$ = "The number of transformations obtained by applying or not applying the transformations $\theta^{-1}_j, \theta^i_j, \theta^{-i}_j$ to each of n hidden neurons" $\times$ "The number of permutations of n hidden neurons obtained by applying $\tau_{j_1 j_2}$" $= 4^n \cdot n! = 2^{2n} \cdot n!$.

References

1. A. Watanabe, N. Yazawa, A. Miyauchi and M. Miyauchi, A method to interpret 3D motions using neural networks, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol.E77-A, No.8, pp.1363-1370 (1994).
2. T. Nitta and T. Furuya, A complex back-propagation learning, *Transactions of Information Processing Society of Japan*, Vol.32, No.10, pp.1319-1329 (1991) (in Japanese).
3. T. Nitta, A complex numbered version of the back-propagation algorithm, *Proc. World Congress on Neural Networks*, Portland, July, Vol.3, pp.576-579 (1993).
4. T. Nitta, An extension of the back-propagation algorithm to complex numbers, *Neural Networks*, Vol.10, No.8, pp.1392-1415 (1997).
5. M. Miura and E. Aiyoshi, Approximation and designing of fractal images by complex neural networks, *Transactions of the Institute of Electrical Engineers of Japan*, C, Vol.123, No.8, pp.1465-1472 (2003) (in Japanese).
6. E. Bingham and A. Hyvarinen, A fast fixed-point algorithm for independent component analysis of complex valued signals, *Int. Journal of Neural Systems*, Vol.10, No.1, pp.1-8 (2000).
7. J.-F. Cardoso and B. H. Laheld, Equivariant adaptive source separation, *IEEE Trans. on Signal Processing*, Vol.44, No.12, pp.3017-3030 (1996).
8. S. Fiori, A. Uncini and F. Piazza, Neural blind separation of complex sources by extended APEX algorithm (EAPEX), *Proc. IEEE Int. Symposium on Circuits and Systems*, Orlando(USA), Vol.V, pp.627-630 (1999).
9. T. Nitta, Orthogonality of decision boundaries in complex-valued neural networks, *Neural Computation*, Vol.16, No.1, pp.73-97 (2004).
10. T. Kim and T. Adali, Approximation by fully-complex mutilayer perceptrons, *Neural Computation*, Vol.15, No.7, pp.1641-1666 (2003).
11. S. Amari, H. Park and T. Ozeki, Singularities affect dynamics of learning in neuromanifolds, *Neural Computation*, Vol.18, No.5, pp.1007-1065 (2006).
12. H. J. Sussmann, Uniqueness of the weights of minimal feedforward nets with a given input-output map, *Neural Networks*, Vol.5, pp.589-593 (1992).
13. K. Fukumizu and S. Amari, Local minima and plateaus in hierarchical structures of multilayer perceptrons, *Neural Networks*, Vol.13, No.3, pp.317-327 (2000).
14. K. Hagiwara, N. Toda and S. Usui, On the problem of applying AIC to determine the structure of a layered feed-forward neural network, *Proc. Int. Joint Conf. on Neural Networks*, Vol.3, pp.2263-2266 (1993).
15. K. Fukumizu, Likelihood ratio of unidentifiable models and multilayer neural networks, *Research Mem-*

orandom No.780, The Institute of Statistical Mathematics (2001).

16. S. Watanabe, Algebraic analysis for non-regular learning machines, in *Advances in Neural Information Processing Systems 12*, eds. S. A. Solla, T. K. Leen and K. Muller, pp.356-362, MIT Press, Cambridge, MA (2000).
17. S. Watanabe, Algebraic analysis for non-identifiable learning machines, *Neural Computation*, Vol.13, No.4, pp.899-933 (2001).
18. T. Nitta, Uniqueness of feedforward complex-valued neural network with a given complex-valued function, in *Knowledge-based Intelligent Information Engineering Systems & Allied Technologies*, eds. N.Baba, L.C.Jain and R.J.Howlett, IOS Press, Tokyo, Part I, pp.550-554 (2001).
19. T. Nitta, The uniqueness theorem for complex-valued neural networks and the redundancy of the parameters, *IEICE Trans. Inf. and Syst.* DII, Vol.J85-D-II, No.5, pp.796-804 (2002) (in Japanese).
(English translation: T. Nitta: The uniqueness theorem for complex-valued neural networks and the redundancy of the parameters, *Systems and Computers in Japan*, Vol.34, Issue 14, pp.54-62 (2003).)
20. A. Hirose, Continuous complex-valued backpropagation learning, *Electronics Letters*, Vol.28, Issue 20, pp.1492-1494 (1992).
21. M. Kobayashi, Redundancy of the parameters of phasor neural networks, *Proc. Workshop on Information-Based Induction Sciences (IBIS 2007)*, Tokyo, Japan, Nov. 5-7 (2007) (in Japanese).
22. G. Thimm and P. Moerland, The interchangeability of learning rate and gain in backpropagation neural networks, *Neural Computation*, Vol.8, No.2, pp.451-460 (1996).
23. D. P. Mandic and J. A. Chambers, Relating the slope of the activation function and the learning rate within a recurrent neural network, *Neural Computation*, Vol.11, No.5, pp.1069-1077 (1999).
24. D. P. Mandic and J. A. Chambers, *Recurrent Nerual Networks for Prediction*, John Wiley & Sons, Ltd (2001).