Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), Vol. 3
Lipo Wang, Jagath C. Rajapakse, Kunihiko Fukushima, Soo-Young Lee, and Xin Yao (Editors)

ON THE CRITICAL POINTS OF THE COMPLEX-VALUED NEURAL NETWORK

Tohru Nitta
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.

ABSTRACT

The properties of the critical points caused by the hierarchical structure of complex-valued neural networks are investigated. If the loss function used is not regular as a complex function, the critical points caused by the hierarchical structure are all saddle points.

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. Nitta et al proposed a multi-layered type complex-valued back-propagation network and demonstrated its characteristics such as an ability to transform geometric figures [5-9].

In the meantime, it was pointed out by Sussmann [10] that there exists the redundancy in the parameters of real-valued neural networks, which is caused by the hierarchical structure of the network. Fukumizu et al. [2] proved the existence of the local minima caused by the hierarchical structure of the real-valued neural networks by using Sussmann's results. Concretely, they proved that a critical point of the model with $H-1$ hidden neurons always gives many critical points of the model H hidden neurons, and these critical points consist of many lines in the parameter space, which could be local minima or saddle points. As is well known, the local minimum of neural networks causes the standstill of learning.

In this paper, the properties of the critical points caused by the hierarchical structure of the complex-valued neural networks are investigated. The main results of this paper are as follows. If the loss function used is not regular as a complex function, the critical points caused by the hierarchical structure form a straight line like those of the real-valued case, whereas they are all saddle points unlike those of the real-valued case. Therefore, we need not to care about the local minima caused by the hierarchical structure when applying the complex-valued neural network to engineering problems.

2. THE COMPLEX-VALUED NEURAL NETWORK

This section describes the 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: $U_n = \sum_m W_{nm} X_m + V_n$, where W_{nm} is the (complex-valued) weight connecting the complex-valued neurons n and m, X_m is the (complex-valued) input signal from the complex-valued neuron m, and V_n is the (complex-valued) threshold value of the 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: $U_n = x + iy = z$, where i denotes $\sqrt{-1}$. The (complex-valued) output signal is defined to be

$$\varphi_C(z) = \tanh(x) + i\tanh(y), \tag{1}$$

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) and is called *hyperbolic tangent*. Note that $-1 < Re[\varphi_C], Im[\varphi_C] < 1$. Note also that $\varphi_C(z)$ is not regular as a complex function.

A complex-valued neural network consists of such complex-valued neurons described above. The network used in the analysis will have 3 layers: $L-H-1$ network. For the sake of simplicity, it is assumed that each threshold parameter of all the hidden neurons is zero, and that the output function ψ_C of the output neuron is linear, that is, $\psi_C(z) = z$ for any $z \in \boldsymbol{C}$ where $\boldsymbol{C}$ denotes the set of complex numbers. For any input pattern $\boldsymbol{x} = (x_1, \cdots, x_L)^T \in \boldsymbol{C}^L$ to the complex-valued neural network where $x_i \in \boldsymbol{C}$ is the input signal to the input neuron i ($1 \le i \le L$) and T denotes transposition, the output value of the output neuron is defined to be

$$f^{(H)}(\boldsymbol{x};\boldsymbol{\theta}^{(H)}) = \sum_{j=1}^{H} \nu_j \varphi_C(\boldsymbol{w}_j^T \boldsymbol{x}) + \nu_0 \in \boldsymbol{C}, \quad (2)$$

where $\boldsymbol{w}_j = (w_{j1}, \cdots, w_{jL})^T \in \boldsymbol{C}^L$ is the weight vector of the hidden neuron j ($w_{ji} \in \boldsymbol{C}$ is the weight between the input neuron i and the hidden neuron j)($1 \le j \le H$), $\nu_j \in \boldsymbol{C}$ is the weight between the hidden neuron j and the output neuron ($1 \le j \le H$), $\nu_0 \in \boldsymbol{C}$ is the threshold of the output neuron, and $\boldsymbol{\theta}^{(H)} = (\nu_0, \nu_1, \cdots, \nu_H, \boldsymbol{w}_1^T, \cdots, \boldsymbol{w}_H^T)^T$ which summarizes all the parameters in one large vector.

Given N complex-valued training data $\{(\boldsymbol{x}^{(\nu)}, y^{(\nu)}) \mid \nu = 1, \cdots, N\}$, we use a complex-valued neural network to realize the relation expressed by the data. The objective of the training is to find the parameters that minimize the error function defined by

$$E_H(\boldsymbol{\theta}^{(H)}) = \sum_{\nu=1}^{N} l(y^{(\nu)}, f^{(H)}(\boldsymbol{x}^{(\nu)}; \boldsymbol{\theta}^{(H)})) \in \boldsymbol{R}, \quad (3)$$

where $l(y,z) : \boldsymbol{C} \times \boldsymbol{C} \longrightarrow \boldsymbol{R}$ is a loss function such that $l(y,z) \ge 0$ and the equality holds if and only if $y = z$. Note that l is not regular as a complex function because it does not take a complex value but a real value. To the author's knowledge, all of the multilayered complex-valued neural networks proposed so far (for example [1,3-9]) employ the mean square error $l(y,z) = (1/2)|y-z|^2$ which takes a real value.

3. CRITICAL POINTS OF THE COMPLEX-VALUED NEURAL NETWORK

This section clarifies the properties of the critical points of the complex-valued neural network described in Section 2.

3.1. Redundancy based on the hierarchical structure

This section makes clear the redundancy based on the hierarchical structure of the complex-valued neural network, that is, the structure of the redundancy of the complex-valued neural network with H hidden neurons for a given set of parameters of the complex-valued neural network with $H-1$ hidden neurons.

Definition 1 *Define*

$$F_H = \{f^{(H)}(\boldsymbol{x};\boldsymbol{\theta}^{(H)}) : \boldsymbol{C}^L \longrightarrow \boldsymbol{C}^1 \mid \boldsymbol{\theta}^{(H)} \in \Theta_H\}, \quad (4)$$

$$\pi_H : \Theta_H \longrightarrow F_H, \qquad \boldsymbol{\theta}^{(H)} \longmapsto f^{(H)}(\boldsymbol{x};\boldsymbol{\theta}^{(H)}), \quad (5)$$

where Θ_H is the set of all the parameters (weights and thresholds) of the $L-H-1$ complex-valued neural network, i.e., $\Theta_H = \boldsymbol{C}^{LH+H+1}$.

F_H is a functional space, the family of all the functions realized by the $L-H-1$ complex-valued neural network. The functional spaces $\{F_H\}_{H=0}^{\infty}$ have a trivial hierarchical structure:

$$F_0 \subset F_1 \subset \cdots \subset F_{H-1} \subset F_H \subset \cdots. \quad (6)$$

π_H gives a complex-valued function for a given set of parameters of the complex-valued neural network with H hidden neurons. Obviously, π_H is not one-to-one: different $\boldsymbol{\theta}^{(H)}$ may give the same input-output function.

Definition 2 *Define*

$$\Omega_H = \{\boldsymbol{\theta}^{(H)} \in \Theta_H \mid \pi_H(\boldsymbol{\theta}^{(H)}) \in i_{H-1}(F_{H-1}(\boldsymbol{\theta}^{(H-1)})), \ \boldsymbol{\theta}^{(H-1)} \in \Theta_{H-1}\}, \quad (7)$$

where $i_{H-1} : F_{H-1} \longrightarrow F_H$, $f \longmapsto i_{H-1}(f) = f$ is the inclusion.

Ω_H is the set of all the parameters $\boldsymbol{\theta}^{(H)}$ that realize the input-output functions of complex-valued neural network with $H-1$ hidden neurons. The following proposition can be easily shown using the results in [7].

Proposition 1 *Ω_H consists of the union of the following complex submanifolds of Θ_H:*

$$\begin{aligned} A_j &= \{\boldsymbol{\theta}^{(H)} \in \Theta_H \mid \nu_j = 0\} \ (1 \le j \le H), && (8) \\ B_j &= \{\boldsymbol{\theta}^{(H)} \in \Theta_H \mid \boldsymbol{w}_j = \boldsymbol{0}\} \ (1 \le j \le H), && (9) \\ C_{j_1 j_2} &= \{\boldsymbol{\theta}^{(H)} \in \Theta_H \mid \boldsymbol{w}_{j_1} = \boldsymbol{w}_{j_2}, \boldsymbol{w}_{j_1} = -\boldsymbol{w}_{j_2}, \\ &\quad \boldsymbol{w}_{j_1} = i\boldsymbol{w}_{j_2} \text{ or } \boldsymbol{w}_{j_1} = -i\boldsymbol{w}_{j_2}\} \\ &\quad (1 \le j_1 < j_2 \le H). && (10) \end{aligned}$$

Proposition 1 shows the structure of the set Ω_H of all the parameters $\boldsymbol{\theta}^{(H)}$ that realize the input-output functions of the complex-valued neural network with $H-1$ hidden neurons.

Next, we examine the structure of the set of all the parameters $\boldsymbol{\theta}^{(H)}$ that realize the input-output function realized by a given set of parameters $\boldsymbol{\theta}^{(H-1)}$.

Definition 3 *Define*

$$\Omega_H(\boldsymbol{\theta}^{(H-1)}) = \pi_H^{-1}(i_{H-1}(f^{(H-1)}(\boldsymbol{x};\boldsymbol{\theta}^{(H-1)}))), \quad (11)$$

where $f^{(H-1)}(\boldsymbol{x},\boldsymbol{\theta}^{(H-1)}) \in F_{H-1} - F_{H-2}$ is a complex function realized by a given set of parameters $\boldsymbol{\theta}^{(H-1)}$ and we use the following notation for its parameters and indexing:

$$f^{(H-1)}(\boldsymbol{x};\boldsymbol{\theta}^{(H-1)}) = \sum_{j=2}^{H} \xi_j \varphi(\boldsymbol{u}_j^T \boldsymbol{x}) + \xi_0. \quad (12)$$

$\Omega_H(\boldsymbol{\theta}^{(H-1)})$ is the set of all the parameters $\boldsymbol{\theta}^{(H)} \in \boldsymbol{\Theta}_H$ that realize a complex function $f^{(H-1)}(\boldsymbol{x}, \boldsymbol{\theta}^{(H-1)})$ realized by a given set of parameters $\boldsymbol{\theta}^{(H-1)}$.

The following proposition can also be easily shown using the results in [7].

Proposition 2 $\Omega_H(\boldsymbol{\theta}^{(H-1)})$ *consists of the ones obtained by transforming* Λ, Ξ *and* Γ *using the transformations in the finite group* $W_{L,H}$ *defined in* [7], *where*

$$\Lambda = \{\boldsymbol{\theta}^{(H)} \in \boldsymbol{\Theta}_H \mid \nu_1 = 0, \nu_0 = \xi_0, \nu_j = \xi_j, \boldsymbol{w}_j = \boldsymbol{u}_j \quad (2 \le j \le H)\}, \tag{13}$$

$$\Xi = \{\boldsymbol{\theta}^{(H)} \in \boldsymbol{\Theta}_H \mid \boldsymbol{w}_1 = \boldsymbol{0}, \nu_1\varphi_C(w_{10}) + \nu_0 = \xi_0, \nu_j = \xi_j, \boldsymbol{w}_j = \boldsymbol{u}_j \quad (2 \le j \le H)\}, \tag{14}$$

$$\Gamma = \{\boldsymbol{\theta}^{(H)} \in \boldsymbol{\Theta}_H \mid \boldsymbol{w}_1 = \boldsymbol{w}_2 = \boldsymbol{u}_2, \nu_0 = \xi_0, \nu_1 + \nu_2 = \xi_2, \nu_j = \xi_j, \boldsymbol{w}_j = \boldsymbol{u}_j \quad (3 \le j \le H)\}. \tag{15}$$

Λ is a complex submanifold of A_1 (eqn (8)) and a complex L-dimensional affine space parallel to the $\boldsymbol{w}_1$-complex plane because only $\boldsymbol{w}_1 \in \boldsymbol{C}^L$ is free (all the other components of $\boldsymbol{\theta}^{(H)}$ are fixed using $\boldsymbol{\theta}^{(H-1)}$). Ξ is a complex submanifold of B_1 (eqn (9)) and a complex 2-dimensional complex submanifold defined by a nonlinear equation $\nu_1\varphi_C(w_{10}) + \nu_0 = \xi_0$ where ν_1, w_{10} and $\nu_0 \in \boldsymbol{C}$ are free, and $\xi_0 \in \boldsymbol{C}$ is fixed. Γ is a complex submanifold of C_{12} (eqn (10)) and a 2-dimensional affine space defined by $\nu_1 + \nu_2 = \xi_2$ where $\nu_1, \nu_2 \in \boldsymbol{C}$ are free, and $\xi_2 \in \boldsymbol{C}$ is fixed.

Here, we define the following canonical embeddings from $\boldsymbol{\Theta}_{H-1}$ to $\boldsymbol{\Theta}_H$, which will be used for the analysis of the critical points in the following sections.

Definition 4 (i) *For any* $\boldsymbol{w} \in \boldsymbol{C}^L$, *define*

$$\begin{aligned}\alpha_{\boldsymbol{w}} : \boldsymbol{\Theta}_{H-1} &\longrightarrow \boldsymbol{\Theta}_H,\\ \boldsymbol{\theta}^{(H-1)} &\longmapsto (\xi_0, 0, \xi_2, \cdots, \xi_H, \boldsymbol{w}^T, \boldsymbol{u}_2^T, \cdots, \boldsymbol{u}_H^T)^T.\end{aligned} \tag{16}$$

(ii) *For any* $\nu, w \in \boldsymbol{C}$, *define*

$$\begin{aligned}\beta_{(\nu,w)} : \boldsymbol{\Theta}_{H-1} &\longrightarrow \boldsymbol{\Theta}_H,\\ \boldsymbol{\theta}^{(H-1)} &\longmapsto (\xi_0 - \nu\varphi_C(w), \nu, \xi_2, \cdots, \xi_H,\\ &\qquad (w, \boldsymbol{0}^T), \boldsymbol{u}_2^T, \cdots, \boldsymbol{u}_H^T)^T.\end{aligned} \tag{17}$$

(iii) *For any* $\lambda \in \boldsymbol{C}$, *define*

$$\begin{aligned}\gamma_\lambda : \boldsymbol{\Theta}_{H-1} &\longrightarrow \boldsymbol{\Theta}_H,\\ \boldsymbol{\theta}^{(H-1)} &\longmapsto (\xi_0, \lambda\xi_2, (1-\lambda)\xi_2, \xi_3, \cdots, \xi_H,\\ &\qquad \boldsymbol{u}_2^T, \boldsymbol{u}_2^T, \boldsymbol{u}_3^T, \cdots, \boldsymbol{u}_H^T)^T.\end{aligned} \tag{18}$$

It is trivial to show that the following proposition holds.

Proposition 3

$$\Lambda = \{\alpha_{\boldsymbol{w}}(\boldsymbol{\theta}^{(H-1)}) \mid \boldsymbol{w} \in \boldsymbol{C}^L\}, \tag{19}$$

$$\Xi = \{\beta_{(\nu,w)}(\boldsymbol{\theta}^{(H-1)}) \mid (\nu, w) \in \boldsymbol{C}^2\}, \tag{20}$$

$$\Gamma = \{\gamma_\lambda(\boldsymbol{\theta}^{(H-1)}) \mid \lambda \in \boldsymbol{C}\}. \tag{21}$$

3.2. Critical points

This section investigates the critical points of the complex-valued neural network.

Generally, the objective of the learning of neural networks is to obtain a global minimum of the error function. If $\boldsymbol{\omega}_*$ is a global minimum of the error function $E(\boldsymbol{\omega})$, the equation $\partial E(\boldsymbol{\omega}_*)/\partial \boldsymbol{\omega} = \boldsymbol{0}$ holds. However, $\partial E(\boldsymbol{\omega}_*)/\partial \boldsymbol{\omega} = \boldsymbol{0}$ does not always assure that $\boldsymbol{\omega}_*$ is a global minimum. The point $\boldsymbol{\omega}_*$ satisfying $\partial E(\boldsymbol{\omega}_*)/\partial \boldsymbol{\omega} = \boldsymbol{0}$ is called a *critical point* of E. There are three types of critical points: a local minimum, a local maximum and a saddle point, which can be identified using Hessian as is well known.

Specifically, we define the critical point and its three types of the complex-valued neural network defined in Section 2.

Definition 5 (i) *A parameter* $\boldsymbol{\theta}^{(H)} = (\theta_1^{(H)}, \cdots, \theta_K^{(H)}) \in \boldsymbol{\Theta}_H$ *is called a critical point of the error function* $E_H(\boldsymbol{\theta}^{(H)})$ *if the following equations hold:*

$$\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Re[\boldsymbol{\theta}^{(H)}]} = \left(\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Re[\theta_1^{(H)}]}, \cdots, \frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Re[\theta_K^{(H)}]}\right)^T = \boldsymbol{0}, \tag{22}$$

$$\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Im[\boldsymbol{\theta}^{(H)}]} = \left(\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Im[\theta_1^{(H)}]}, \cdots, \frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Im[\theta_K^{(H)}]}\right)^T = \boldsymbol{0}, \tag{23}$$

where $K = LH + H + 1$ *is the number of the parameters of the complex-valued neural network.*

(ii) *A critical point* $\hat{\boldsymbol{\theta}}^{(H)} \in \boldsymbol{\Theta}_H$ *is called a local minimum (maximum) if there exists a neighborhood around* $\hat{\boldsymbol{\theta}}^{(H)}$ *such that for any point* $\boldsymbol{\theta}^{(H)}$ *in the neighborhood* $E_H(\boldsymbol{\theta}^{(H)}) \ge E_H(\hat{\boldsymbol{\theta}}^{(H)})$ $(E_H(\boldsymbol{\theta}^{(H)}) \le E_H(\hat{\boldsymbol{\theta}}^{(H)}))$ *holds, and called a saddle if it is neither a local minimum nor a local maximum.*

The next proposition can be easily shown by simple calculations.

Proposition 4 *Let* $\boldsymbol{\theta}_*^{(H-1)} = (\xi_{0_*}, \xi_{2_*}, \cdots, \xi_{H_*}, \boldsymbol{u}_{2_*}^T, \cdots, \boldsymbol{u}_{H_*}^T)^T \in \boldsymbol{\Theta}_{H-1} - \boldsymbol{\Theta}_{H-2}$ *be a critical point of*

E_{H-1}. *Then, for any* $2 \leq j \leq H$, *the following equations hold:*

$$\frac{\partial E_{H-1}(\boldsymbol{\theta}_*^{(H-1)})}{\partial Re[\xi_0]} = \frac{\partial E_{H-1}(\boldsymbol{\theta}_*^{(H-1)})}{\partial Im[\xi_0]} = 0, \tag{24}$$

$$\frac{\partial E_{H-1}(\boldsymbol{\theta}_*^{(H-1)})}{\partial Re[\xi_j]} = \frac{\partial E_{H-1}(\boldsymbol{\theta}_*^{(H-1)})}{\partial Im[\xi_j]} = 0, \tag{25}$$

$$\begin{aligned}
&\frac{\partial E_{H-1}(\boldsymbol{\theta}_*^{(H-1)})}{\partial Re[\boldsymbol{u}_j]} = \\
&\sum_{\nu=1}^{N}\left[\frac{\partial l}{\partial z}\left(y^{(\nu)}, f^{(H-1)}(\boldsymbol{x}^{(\nu)}, \boldsymbol{\theta}_*^{(H-1)})\right)\right. \\
&\xi_{j*}\left\{\frac{\partial \varphi_C(\boldsymbol{u}_{j*}^T\boldsymbol{x}^{(\nu)})}{\partial z}\boldsymbol{x}^{(\nu)T} + \frac{\partial \varphi_C(\boldsymbol{u}_{j*}^T\boldsymbol{x}^{(\nu)})}{\partial \bar{z}}\overline{\boldsymbol{x}^{(\nu)}}^T\right\} \\
&\quad + \frac{\partial l}{\partial \bar{z}}\left(y^{(\nu)}, f^{(H-1)}(\boldsymbol{x}^{(\nu)}, \boldsymbol{\theta}_*^{(H-1)})\right) \\
&\left.\overline{\xi_{j*}}\left\{\frac{\partial \overline{\varphi_C(\boldsymbol{u}_{j*}^T\boldsymbol{x}^{(\nu)})}}{\partial z}\boldsymbol{x}^{(\nu)T} + \frac{\partial \overline{\varphi_C(\boldsymbol{u}_{j*}^T\boldsymbol{x}^{(\nu)})}}{\partial \bar{z}}\overline{\boldsymbol{x}^{(\nu)}}^T\right\}\right] \\
&\quad = \boldsymbol{0},
\end{aligned} \tag{26}$$

$$\begin{aligned}
&\frac{\partial E_{H-1}(\boldsymbol{\theta}_*^{(H-1)})}{\partial Im[\boldsymbol{u}_j]} = \\
&i\sum_{\nu=1}^{N}\left[\frac{\partial l}{\partial z}\left(y^{(\nu)}, f^{(H-1)}(\boldsymbol{x}^{(\nu)}, \boldsymbol{\theta}_*^{(H-1)})\right)\right. \\
&\xi_{j*}\left\{\frac{\partial \varphi_C(\boldsymbol{u}_{j*}^T\boldsymbol{x}^{(\nu)})}{\partial z}\boldsymbol{x}^{(\nu)T} - \frac{\partial \varphi_C(\boldsymbol{u}_{j*}^T\boldsymbol{x}^{(\nu)})}{\partial \bar{z}}\overline{\boldsymbol{x}^{(\nu)}}^T\right\} \\
&\quad + \frac{\partial l}{\partial \bar{z}}\left(y^{(\nu)}, f^{(H-1)}(\boldsymbol{x}^{(\nu)}, \boldsymbol{\theta}_*^{(H-1)})\right) \\
&\left.\overline{\xi_{j*}}\left\{\frac{\partial \overline{\varphi_C(\boldsymbol{u}_{j*}^T\boldsymbol{x}^{(\nu)})}}{\partial z}\boldsymbol{x}^{(\nu)T} - \frac{\partial \overline{\varphi_C(\boldsymbol{u}_{j*}^T\boldsymbol{x}^{(\nu)})}}{\partial \bar{z}}\overline{\boldsymbol{x}^{(\nu)}}^T\right\}\right] \\
&\quad = \boldsymbol{0}.
\end{aligned} \tag{27}$$

The next theorem shows the existence of the critical points of the complex-valued neural network.

Theorem 1 *Let* $\boldsymbol{\theta}_*^{(H-1)} = (\xi_{0*}, \xi_{2*}, \cdots, \xi_{H*}, \boldsymbol{u}_{2*}^T, \cdots, \boldsymbol{u}_{H*}^T)^T \in \Theta_{H-1} - \Theta_{H-2}$ *be a critical point of* E_{H-1}. *Let* γ_λ *be as in eqn* (18). *Then,*
(i) *The point* $\gamma_\lambda(\boldsymbol{\theta}_*^{(H-1)})$ *is a critical point of* E_H *for any* $\lambda \in \boldsymbol{R}$.
(ii) *The point* $\gamma_\lambda(\boldsymbol{\theta}_*^{(H-1)})$ *is not always a critical point of* E_H *for any* $\lambda \in \boldsymbol{C}$ *s.t.* $Im[\lambda] \neq 0$.

Proof. Let $\boldsymbol{\theta}^{(H)} = \gamma_\lambda(\boldsymbol{\theta}_*^{(H-1)})$ for any $\lambda \in \boldsymbol{C}$. First, we can easily see from Proposition 4 and the equation $f^{(H)}(\boldsymbol{x}, \boldsymbol{\theta}^{(H)}) = f^{(H-1)}(\boldsymbol{x}, \boldsymbol{\theta}_*^{(H-1)})$ that the following equations hold: for any $0 \leq j \leq H$

$$\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Re[\nu_j]} = \frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Im[\nu_j]} = 0, \tag{28}$$

and for any $3 \leq j \leq H$

$$\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Re[\boldsymbol{w}_j]} = \frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Im[\boldsymbol{w}_j]} = \boldsymbol{0}. \tag{29}$$

However, in the case of $j = 1$, since

$$\begin{aligned}
&\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Re[\boldsymbol{w}_1]} = \\
&\sum_{\nu=1}^{N}\left[\frac{\partial l}{\partial z}\left(y^{(\nu)}, f^{(H-1)}(\boldsymbol{x}^{(\nu)}, \boldsymbol{\theta}_*^{(H-1)})\right)\right. \\
&\lambda\xi_{2*}\left\{\frac{\partial \varphi_C(\boldsymbol{u}_{2*}^T\boldsymbol{x}^{(\nu)})}{\partial z}\boldsymbol{x}^{(\nu)T} + \frac{\partial \varphi_C(\boldsymbol{u}_{2*}^T\boldsymbol{x}^{(\nu)})}{\partial \bar{z}}\overline{\boldsymbol{x}^{(\nu)}}^T\right\} \\
&\quad + \frac{\partial l}{\partial \bar{z}}\left(y^{(\nu)}, f^{(H-1)}(\boldsymbol{x}^{(\nu)}, \boldsymbol{\theta}_*^{(H-1)})\right) \\
&\left.\overline{\lambda\xi_{2*}}\left\{\frac{\partial \overline{\varphi_C(\boldsymbol{u}_{2*}^T\boldsymbol{x}^{(\nu)})}}{\partial z}\boldsymbol{x}^{(\nu)T} + \frac{\partial \overline{\varphi_C(\boldsymbol{u}_{2*}^T\boldsymbol{x}^{(\nu)})}}{\partial \bar{z}}\overline{\boldsymbol{x}^{(\nu)}}^T\right\}\right],
\end{aligned} \tag{30}$$

$$\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Re[\boldsymbol{w}_1]}\begin{cases} = 0 & (\text{if } \lambda \in \boldsymbol{R}) \ (\text{from eqn (26)}) \\ \neq 0 & (\text{if } Im[\lambda] \neq 0, \text{i.e.}, \lambda \notin \boldsymbol{R}). \end{cases} \tag{31}$$

Similarly,

$$\frac{\partial E_H(\boldsymbol{\theta}^{(H)})}{\partial Re[\boldsymbol{w}_2]}\begin{cases} = 0 & (\text{if } \lambda \in \boldsymbol{R}) \ (\text{from eqn (26)}) \\ \neq 0 & (\text{if } Im[\lambda] \neq 0, \text{i.e.}, \lambda \notin \boldsymbol{R}). \end{cases} \tag{32}$$

The similar equations hold for the papameters $Im[\boldsymbol{w}_j] (j = 1, 2)$ because of eqn (27). This completes the proof. (Q.E.D.)

As described in Section 2, the loss function $l : \boldsymbol{C} \times \boldsymbol{C} \longrightarrow \boldsymbol{R}$ is not *regular* as a complex function. This is the reason why the point $\gamma_\lambda(\boldsymbol{\theta}_*^{(H-1)})$ of Theorem 1 (ii) is not always a critical point.

Theorem 2 *The critical point of Theorem 1 (i) consists of a straight line in the 2-dimensional affine space defined by* $\nu_1 + \nu_2 = \xi_{2*}$ *if we move* $\lambda \in \boldsymbol{R}$.

Proof. Since $\nu_1 = \lambda\xi_{2*}$ and $\nu_2 = (1 - \lambda)\xi_{2*}$, the following equation holds:

$$\begin{bmatrix} Re[\nu_1] \\ Im[\nu_1] \\ Re[\nu_2] \\ Im[\nu_2] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ Re[\xi_{2*}] \\ Im[\xi_{2*}] \end{bmatrix} + \lambda \begin{bmatrix} Re[\xi_{2*}] \\ Im[\xi_{2*}] \\ -Re[\xi_{2*}] \\ -Im[\xi_{2*}] \end{bmatrix}. \tag{33}$$

(Q.E.D.)

Theorem 3 *There exist many critical points each of which forms a straight line in the complex-valued neural network.*

Proof. If $\boldsymbol{\theta}$ is a critical point of Theorem 1 (i), so is $T(\boldsymbol{\theta})$ for any transformation $T \in W_{L,H}$, where $W_{L,H}$ is a finite group defined in [7]. (Q.E.D.)

It should be noted here that the critical points described in this section are only the ones caused by the hierarchical structures of the complex-valued neural network, especially the ones based on the embedding γ_λ defined by eqn (18).

3.3. Saddle points

This section investigates the property of the critical points of Theorem 1 (i).

Theorem 4 *The critical points of Theorem 1 (i) are all saddle points.*

Proof. We need the following lemma which Fukumizu et al. proved in [2].

Lemma 1 *Let $E(\boldsymbol{\theta})$ be a function of class C^1, and $\boldsymbol{\theta}_*$ be a critical point of $E(\boldsymbol{\theta})$. If in any neighborhood of $\boldsymbol{\theta}_*$ there exists a point $\boldsymbol{\theta}$ such that $E(\boldsymbol{\theta}) = E(\boldsymbol{\theta}_*)$ and $\partial E(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \neq \mathbf{0}$, then $\boldsymbol{\theta}_*$ is a saddle point.*

Back to the proof of Theorem 4, from Theorem 2, the set of the critical points $\{\gamma_\lambda(\boldsymbol{\theta}_*^{(H-1)}) \mid \lambda \in \boldsymbol{R}\}$ is a straight line in the 2-dimensional affine space defined by $\nu_1 + \nu_2 = \xi_{2_*}$. And it is obvious that the error function E_H takes the same value in the 2-dimensional affine space defined by $\nu_1 + \nu_2 = \xi_{2_*}$. Here, for any $w \in \boldsymbol{C}$ s.t. $w \notin \boldsymbol{R}$, let $\hat{\boldsymbol{\theta}}^{(H)} = (\xi_{0_*}, w\xi_{2_*}, (1-w)\xi_{2_*}, \xi_{3_*}, \cdots, \xi_{H_*}, \boldsymbol{u}_{2_*}^T, \boldsymbol{u}_{2_*}^T, \boldsymbol{u}_{3_*}^T, \cdots, \boldsymbol{u}_{H_*}^T)^T$. Obviously, $\hat{\boldsymbol{\theta}}^{(H)} \in \Theta_H$ belongs to the 2-dimensional affine space defined by $\nu_1 + \nu_2 = \xi_{2_*}$, and is, however, not on the straight line which the set of the critical points $\{\gamma_\lambda(\boldsymbol{\theta}_*^{(H-1)}) \mid \lambda \in \boldsymbol{R}\}$ forms. Furthermore, we can easily see that $\partial E_H(\hat{\boldsymbol{\theta}}^{(H)})/\partial Re[\boldsymbol{w}_1] \neq \mathbf{0}$ generally. And $\hat{\boldsymbol{\theta}}^{(H)}$ can belong to the any neighborhood of the straight line $\{\gamma_\lambda(\boldsymbol{\theta}_*^{(H-1)}) \mid \lambda \in \boldsymbol{R}\}$. Therefore, from Lemma 1, the critical points of Theorem 1 (i) are all saddle points. (Q.E.D.)

Theorem 5 *There exist many saddle points each of which forms a straight line in the complex-valued neural network.*

Proof. If $\boldsymbol{\theta}$ is a saddle point of Theorem 4, so is $T(\boldsymbol{\theta})$ for any transformation $T \in W_{L,H}$, where $W_{L,H}$ is a finite group defined in [7]. (Q.E.D.)

In the case of the complex-valued neural network, the critical points with respect to the embedding γ_λ are all saddle points, unlike the real-valued neural network in which the critical point with respect to the corresponding embedding γ_λ can be a local minimum or a saddle point [2].

4. CONCLUSIONS

The properties of the critical points caused by the hierarchical structure of the complex-valued neural networks were investigated. The critical points related to the other embeddings $\alpha_{\boldsymbol{w}}$ and $\beta_{(\nu,w)}$ and their properties will be reported in a future paper.

Acknowledgements

We wish to thank the members of the Mathematical Neuroinformatics Group and the reviewers for their helpful suggestions.

5. REFERENCES

[1] N. Benvenuto and F. Piazza, "On the Complex Backpropagation Algorithm", *IEEE Trans. Signal Processing*, Vol.40, No.4, pp.967-969 (1992).

[2] 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).

[3] G. M. Georgiou and C. Koutsougeras, "Complex Domain Backpropagation", *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Process.*, Vol.39, No.5, pp.330-334 (1992).

[4] M. S. Kim and C. C. Guest, "Modification of Backpropagation Networks for Complex-valued Signal Processing in Frequency Domain," in *Proc. IJCNN*, Vol.3, June, pp.27-31 (1990).

[5] T. Nitta et al., "A Complex Back-Propagation Learning", *Trans. of Inf. Proc. Soc. of Japan*, Vol.32, No.10, pp.1319-1329 (in Japanese) (1991).

[6] T. Nitta, "An Extension of the Back-Propagation Algorithm to Complex Numbers", *Neural Networks*, Vol.10, No.8, pp.1392-1415 (1997).

[7] T. Nitta, "Uniqueness of Feedforward Complex-valued Neural Network with a Given Complex-valued Function" in *Proc. KES2001*, Part I, pp.550-554 (2001).

[8] T. Nitta, "On the Inherent Property of the Decision Boundary in Complex-valued Neural Networks", *Neurocomputing* (to appear).

[9] T. Nitta, "Redundancy of the Parameters of the Complex-valued Neural Network", *Neurocomputing* (to appear).

[10] 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).