Neural Networks with Comparatively Few Critical Points

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

Abstract

A critical point is a point on which the derivatives of an error function are all zero. It has been shown in the literatures that the critical points caused by the hierarchical structure of the real-valued neural network could be local minima or saddle points, whereas most of the critical points caused by the hierarchical structure are saddle points in the case of complex-valued neural networks. Several studies have demonstrated that that kind of singularity has a negative effect on learning dynamics in neural networks. In this paper, we will demonstrate via some examples that the decomposition of high-dimensional NNs into real-valued NNs equivalent to the original NNs yields the NNs that do not have critical points based on the hierarchical structure.

© 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of KES International

Keywords: critical point; singular point; redundancy; complex number; hyperbolic number; quaternion

1. Introduction

It has been revealed that a singular point affects the training dynamics of a learning model and engenders stagnation of training [1-3]. Learning models with a hierarchical structure or symmetry about exchange of weights, such as a hierarchical neural network (NN) and Gaussian mixture model, usually have a singular point. In a learning process of hierarchical NNs, a parameter $\mathbf{w}^*$ satisfying $\frac{\partial E(\mathbf{w}^*)}{\partial \mathbf{w}} = \mathbf{0}$ is referred to as the critical point of error function $E(\mathbf{w})$ which represents an error between the actual output of the NN and the desired output, and is often defined as $\sum_{k=1}^{N}(o_k-t_k)^2$ where $o_k$ is the actual output and $t_k$ the desired output of the NN. A critical point can be a local minimum, a local maximum, or a saddle point. The critical point is a kind of singular points and causes a standstill of learning.

* Corresponding author
E-mail address: tohru-nitta@aist.go.jp
Fukumizu and Amari have proved the existence of a local minimum resulting from a hierarchical structure in a real-valued NN (usual NN handling real-valued signals) [4]. They have demonstrated that critical points in a three-layer real-valued NN with $H-1$ hidden neurons behave as critical points in a three-layer real-valued NN with $H$ hidden neurons, and that they are local minima or saddle points. Such a critical point is called the critical point based on a hierarchical structure in this paper.

Complex-valued NN also has the critical point based on a hierarchical structure. A complex-valued NN extends (real-valued) parameters such as weight and threshold value in a usual NN to complex numbers. It is suitable for information processing of complex-valued data and two-dimensional data, and is applied to communications, image-processing, biologic information processing, land-mine detection, wind prediction, independent component analysis (ICA), etc., for instance [5-6]. Reportedly, a critical point in a three-layer complex-valued NN also behaves in the same manner as that in a three-layer real-valued NN [7-8]: critical points in a three-layer complex-valued NN with $H-1$ hidden neurons turn into critical points in a three-layer complex-valued NN with $H$ neurons, which are saddle points (except for cases meeting rare conditions).

This paper presents an attempt to implement an NN having no critical point based on a hierarchical structure.

2. Construction of Neural Networks

In this chapter, we demonstrate that NNs having no critical point based on a hierarchical structure can be constructed by decomposing a high-dimensional NN into a real-valued NN.

2.1. Construction of 2-2-2 real-valued NN using a complex-valued NN

A 2-2-2 real-valued NN having no critical point based on a hierarchical structure is constructed from a 1-1-1 complex-valued NN.

Consider a 1-1-1 complex-valued NN (called NET 1 here). We will use $a_i + ia_2 \in \mathbb{C}$ for the weight between the input neuron and the hidden neuron, $c_1 + ic_2 \in \mathbb{C}$ for the weight between the hidden neuron and the output neuron, $b_1 + ib_2 \in \mathbb{C}$ for the threshold of the hidden neuron, and $d_1 + id_2 \in \mathbb{C}$ for the threshold of the output neuron where $i$ denotes $\sqrt{-1}$ and $\mathbb{C}$ denotes the set of complex numbers. We assume that $a_i + ia_2 \neq 0$ and $c_1 + ic_2 \neq 0$. Let $x + iy \in \mathbb{C}$ denote the input signal, and $X + iY \in \mathbb{C}$ the output signal, respectively. We will use the activation functions defined by the following equations:

$$f_c(z) = \tanh(z^x) + i \tanh(z^y), \quad z = z^x + iz^y \in \mathbb{C} \quad (1)$$

for the hidden neuron, and

![Diagram](image_url)

Fig. 1. NET 2, a 2-2-2 real-valued NN equivalent to the NET 1. Restriction applies between weights.
Proposition 1

The NET 2 has no critical point based on a hierarchical structure. □

Proof) Assume a 2-1-2 real-valued NN obtained by removing the hidden neuron 1 from the NET 2 (called NET 3 here) (Fig. 2). Also assume that the learning parameter of the NET 3 is a critical point that implements mapping $F_1(x, y)$. It is necessary to realize any one of the following three conditions for implementing the same mapping $F_1$ by appending once-removed hidden neuron 1 to the NET 3 again [4].

1. Case 1
A weight vector between hidden neuron 1 appended and the two output neurons is 0.

2. Case 2
A weight vector between hidden neuron 1 appended and the two input neurons is 0.

3. Case 3
For the weight vector $w_1$ between hidden neuron 1 appended and the two input neurons, and the weight vector $w_2$ between hidden neuron 2 and the two input neurons, $a_1 = a_2 = 0$ must hold in this case, but this violates the assumption $a_1 + ia_2 \neq 0$.

Therefore, mapping $F_1$ cannot be implemented by the NET 3 with the original hidden neuron 1 appended and having the weight structure of the NET 2.

The above illustrates a case in which hidden neuron 1 is removed, but removal of the hidden neuron 2 engenders the same conclusion. Consequently, the NET 2 has no critical point based on a hierarchical structure.

Note that the choice of activation functions is not essential for the proof. For the reference, the activation functions of the NET 2 are $(\tanh(u), u \in \mathbb{R})$ for the hidden neuron and $(\tanh'(z), z \in \mathbb{C})$ for the output neuron where $\mathbb{R}$ is the set of real numbers. □

2.2. Construction of 2-2-2 real-valued NN using 1-1-1 hyperbolic NN

In this section, a 2-2-2 real-valued NN having no critical point based on a hierarchical structure is constructed from a 1-1-1 hyperbolic NN.

Hyperbolic NN is an extension of usual real-valued NNs to two dimensions [10-12]. It is based on hyperbolic numbers which are a counterpart to complex numbers. Hyperbolic numbers are numbers of the form $w = a + ub$ where $a, b \in \mathbb{R}$ and $u$ is called unipotent which has the algebraic property that $u^2 = 1$ but $u^2 = -1$ [13]. It follows that multiplication in $\mathbb{H}$ is defined by $(a + ub)(c + ud) = (ac + bd) + u(ad + bc)$ where $\mathbb{H}$ denotes the set of hyperbolic numbers. The hyperbolic modulus of $w = a + ub \in \mathbb{H}$ is defined by $|w| = \sqrt{a^2 - b^2}$. If $|a| = |b|$ then $w$ has zero modulus. Hence $\mathbb{H}$ contains divisors of zero.

Consider a 1-1-1 hyperbolic NN (called NET 4 here). We will use $a_1 + ua_2 \in \mathbb{H}$ for the weight between the input neuron and the hidden neuron, $c_1 + uc_2 \in \mathbb{H}$ for the weight between the hidden neuron and the output neuron, $b_1 + ib_2 \in \mathbb{C}$ for the threshold of the hidden neuron, and $d_1 + ud_2 \in \mathbb{H}$ for the threshold of the output neuron. We assume that $a_1 + ua_2 \neq 0$ and $c_1 + uc_2 \neq 0$. Let $x + uy \in \mathbb{H}$ denote the input signal, and $X + uY \in \mathbb{H}$ the output signal, respectively. We will use the activation functions defined by the following equations:

$$f_{iH}(z) = \tanh(z^R) + u \tanh(z'), \quad z = z^R + uz' \in \mathbb{H}$$

for the hidden neuron, and

$$g_{iH}(z) = z, \quad z \in \mathbb{H}$$

for the output neuron. This 1-1-1 complex-valued NN is apparently equivalent to a 2-2-2 real-valued NN (called NET 2 here) shown in Fig.1 [9].
Fig. 2. NET 3, a 2-1-2 real-valued NN obtained by removing the hidden neuron 1 from the NET 2.

Fig. 3. NET 5, a 2-2-2 real-valued NN equivalent to the NET 4. Restriction applies between weights.

for the output neuron. This 1-1-1 hyperbolic NN is apparently equivalent to a 2-2-2 real-valued NN (called NET 5 here) shown in Fig. 3.

**Proposition 2**
The NET 5 has no critical point based on a hierarchical structure. □

(Proof) This can be proved by the same way as Proposition 1. □

2.3. Construction of 4-4-4 real-valued NN

A 4-4-4 real-valued NN having no critical point based on a hierarchical structure is constructed from a 1-1-1 quaternionic NN. The quaternionic NN [14-19] is an extension of the classical real-valued neural network to quaternions, whose weights, threshold values, input and output signals are all quaternions where a quaternion is a four-dimensional number and was invented by W. R. Hamilton in 1843[20].

Consider a 1-1-1 quaternionic NN (called NET 6 here). Let the weight between input neuron and hidden neuron be $A = a_1 + ia_2 + ja_3 + ka_4 \in \mathbb{Q}$, and weight between hidden neuron and output neuron be $C = c_1 + ic_2 + jc_3 + kc_4 \in \mathbb{Q}$ where $\mathbb{Q}$ represents a set of the quaternions. We assume that $A \neq 0$ and $C \neq 0$. Let also $B = b_1 + ib_2 + jb_3 + kb_4 \in \mathbb{Q}$ denote the threshold of the hidden neuron, $D = d_1 + id_2 + jd_3 + kd_4 \in \mathbb{Q}$ the threshold of the output neuron, $I = v + iw + jx + ky \in \mathbb{Q}$ the input signal, and $O = V + iW + jX + kY \in \mathbb{Q}$ the output signal, respectively. We will use the activation functions defined by the following equations:

$$f_Q(u) = \tanh(u_1) + i \tanh(u_2) + j \tanh(u_3) + k \tanh(u_4), \quad u = u_1 + iu_2 + ju_3 + ku_4 \in \mathbb{Q}$$

(5)

for the hidden neuron, and
Figure 4 NET 7, a 4-4-4 real-valued NN equivalent to the NET 6. Restriction applies between weights.

\[ g_Q(u) = u, \quad u \in \mathbb{Q} \]  

for the output neuron. Because a quaternion is non-commutative about multiplication, the computational result varies with the multiplication sequence of an input value and weight: \( I A \neq AI \). Accordingly, quaternion neurons of two kinds exist: a normal quaternion neuron (computing \( AI \)) and an inverse quaternion neuron (computing \( IA \)) [14][17]. This paper specifically addresses a quaternionic NN that comprises only inverse quaternionic neurons as an example.

The NET 6 is apparently equivalent to a 4-4-4 real-valued NN (called NET 7 here) shown in Fig.4.

**Proposition 3**
The NET 7 has no critical point based on a hierarchical structure. □

(Proof) Assume a 4-3-4 real-valued NN obtained by removing the hidden neuron 1 from the NET 7 (called NET 8 here). Also assume that the learning parameter of the NET 8 is a critical point that implements mapping \( F_2(v, w, x, y) \). It is necessary to realize any one of the following three conditions for implementing the same mapping \( F_2 \) by appending once-removed hidden neuron 1 to the NET 8 again [4].

1. Case 1
A weight vector between hidden neuron 1 appended and the four output neurons is \( 0 \). 
\( C = 0 \) must hold in this case, but this violates the assumption \( C \neq 0 \).

2. Case 2
A weight vector between hidden neuron 1 appended and the four input neurons is \( 0 \). 
\( A = 0 \) must hold in this case, but this violates the assumption \( A \neq 0 \).
3. Case 3
Let $\mathbf{w}_j$ denote the weight vector between hidden neuron $j$ and the four input neurons for any $1 \leq j \leq 4$ where the hidden neuron 1 is the appended one. Then, there exists some $2 \leq j \leq 4$ such that $\mathbf{w}_i = \mathbf{w}_j$ or $\mathbf{w}_i = -\mathbf{w}_j$. $A = 0$ must hold in this case, but this violates the assumption $A \neq 0$.

Therefore, mapping $F_2$ cannot be implemented by the NET 8 with the original hidden neuron 1 appended and having the weight structure of the NET 7. The above illustrates a case in which hidden neuron 1 is removed, but removal of the hidden neuron $j$ engenders the same conclusion ($2 \leq j \leq 4$). Consequently, the NET 7 has no critical point based on a hierarchical structure.

3. Discussion

In Sections 2.1 and 2.2, the two 2-2-2 real-valued NNs with the same network structure from a 1-1-1 complex-valued NN and a 1-1-1 hyperbolic NN have been constructed. However, the weight structure of each 2-2-2 real-valued NN obtained is different. Both the complex-valued NN and the hyperbolic NN belong to the two-dimensional Clifford NN [21]. The two-dimensional Clifford NN consists merely of these two types. More generally, as a $2^n$-dimensional Clifford NN has $n+1$ kinds of different Clifford NNs, it is supposed that $n+1$ kinds of $2^n-2^n-2^n$ real-valued NNs corresponding to each of them can be constructed by decomposing them. That is, $n+1$ kinds of real-valued NNs having no critical point based on a hierarchical structure with the identical $2^n-2^n-2^n$ network structure, but different weight can be obtained.

4. Conclusion

This paper proposes a construction process of a NN having no critical point based on a hierarchical structure: we demonstrate that real-valued NN having no critical point based on a hierarchical structure can be constructed by decomposing a high-dimensional NN into equivalent real-valued NN. Concretely, the following three cases are shown: (a) A 2-2-2 real-valued NN is constructed from a 1-1-1 complex-valued NN. (b) A 2-2-2 real-valued NN is constructed from a 1-1-1 hyperbolic NN. (c) A 4-4-4 real-valued NN is constructed from a 1-1-1 quaternionic NN. Those NNs described above do not suffer from a negative effect by singular points during learning comparatively because they do not have critical points based on a hierarchical structure. Although we gave only three examples in this paper, we believe that we could present a general process of a NN having no critical point based on a hierarchical structure.

The author expects to address the following issues in future studies. (a) Although quaternionic NN that comprises only inverse quaternary neurons are treated in this paper, the case with normal quaternary neurons shall be considered. (b) A $2^s$-dimensional Clifford NN having no critical point based on a hierarchical structure shall be produced by decomposing a general $2^n$-dimensional Clifford NN into equivalent Clifford NNs of $2^s$ dimensions ($s < n$).

Acknowledgements

The author would like to give special thanks to the members of the Mathematical Neuroinformatics Group and the anonymous reviewers for valuable comments.
References