Complex-Valued and Quaternion Neural Networks
to Deal with Complex Amplitude and Polarization:
Their Strength in Electronics

Akira Hirose*
* The University of Tokyo, Tokyo, Japan
E-mail: ahirose at ee.t.u-tokyo.ac.jp

Abstract—This paper discusses the advantages in complex-valued and quaternion neural networks (CVNNs and QNNs) by presenting important concepts as well as applications in electronic engineering. Essentially, CVNNs and QNNs have excellent generalization ability in the learning of two- and three-dimensional rotation and scaling, respectively. Its origin is the sparsity in themselves as complex and quaternion numbers. The ability is useful in particular in adaptive processing of waves consisting of complex amplitude (amplitude and phase) and/or polarization. We consider an example in communications where we deal with channel prediction to discuss the superiority in CVNNs and its basics.

I. INTRODUCTION

Neural network research and development have been promoted forward by the cooperation of mathematics, physics, physiology and psychology to elucidate the human brain mechanisms as well as to utilize its findings in engineering so that we can construct systems which provide us with great functionality beyond human beings in specific aspects [2]–[8]. Complex-valued neural networks (CVNNs) and quaternion neural networks (QNNs) are constructed by, and dealing with, complex and quaternion numbers, respectively, to realize such super-human systems.

Most CVNNs and QNNs aim at engineering applications rather than human physiology or psychology [9]–[12]. In electrical and electronic engineering, CVNNs are the networks dealing with complex amplitude. That is, they are coherent neural networks, that treat phase information in a consistent manner. They are very powerful in phase-sensitive physics and systems such as satellite/air-borne synthetic aperture radar (SAR) [13]–[16], ground penetrating radar (GPR) [17]–[20], quantum computing [21], [22], and lightwave computing [23], [24]. Analyses of the performance and dynamics in such systems are also very active in various fields such as communication systems [25] and related mathematics in complex manifold [26], [27]. Similarly, QNNs are widely effective in polarization processing when we represent its state on/in the Poincare sphere because of the high generalization ability in three-dimensional space [28]–[30]. Many ideas have been proposed on supervised land classification [31]–[34], unsupervised classification [35], [36], and also self-organizing codebook generation [37]. Moreover, device applications are also within the CVNN applications. For example, a reservoir computing device using spin waves without hard wiring has also been proposed [38]–[42].

We can observe these trends from an inverse-directional sight as follows. Adaptive signal processing based on learning and self-organization has a long history in such fields as adaptive antennas and inverse problems [43]. In linear systems, the complex least mean square (LMS) algorithm was proposed in 1975 [44] and used in various steepest-descent and conjugate-gradient systems [45]. CVNN is a more generalized framework by introducing nonlinearity, layers and various connection forms. Recent special issue on CVNNs [46] also focuses on such aspects of CVNNs in their theories [47], [48] and applications [49]–[52].

People just using neural networks as convenient tools may sometimes say, "a neural network solves problem, but it is a black box that presents nothing about the logical deduction or reasoning." It is true that a neural network is not a "logic" machine, but actually it presents the reasons. In reality, there exist many networks which analyze problems and/or extract features, namely, deep-learning networks based on classical hourglass-structure neural networks [53], [54], and convolutional neural networks including the Neocognitron [55], [56]. Neural networks reflect the external world in their connections. Then, it is also possible for us to prepare a network suitable for respective tasks to be dealt with. A good example is adaptive antennas. CVNN is a network incorporating the wave physics. It is also a combination of things (physics) of electromagnetic wave propagation and scattering and things (maths, affairs) of adaptive and statistical information processing.

CVNNs show the merits in, for example, earth observation from space, several hundred kilometers above land, by using microwave transmitted from a satellite and processing the scattered wave data with the synthetic aperture process, to obtain earth surface information included basically in the amplitude, phase and polarization [57]. The raw data often suffer from the distortion generated by propagation, diffraction, refraction, interference and spatial and temporal discreteness in observation. Restoration of the real information included in the distorted data requires good estimation and prediction methods. Here the CVNNs, supposing wave physics, play an
important role [58], [59]. Their estimation and prediction are the result of the combination of mathematics of adaptation and physics of waves. The application fields are expanding continuously. The precise and global observation of earth surface forms and motions will realize dense and quasi-real-time monitoring of volcano activities, earthquake damages, changes in glaciers and pole ices as well as forest biomass in relation to global warming, agricultural product based on polarization changes, and so on and so forth. All of them are related to the key issues to be solved in the present society, namely, sustainable development goals (SDGs).

This paper discusses the essence of CVNNs by focusing on the generalization ability. Then we check the merit briefly in communication channel prediction. In the presentation, we discuss also QNNs as well.

II. GENERALIZATION ABILITY

Generalization ability is the most important concept in neural networks and other adaptive signal processing. We review and discuss the key points in the generalization ability of CVNNs [58].

A. Sparsity and degree of freedom

In general, a neural network realizes its desired function by adjusting the connection weights among neurons depending on fed signals and environments. The adjustment is preformed through supervised learning, unsupervised learning or reinforcement learning. The more the weights exist, the higher the degree of freedom is.

Fig. 1(a) illustrates a function-approximation task which a neural network is expected to learn by supervised learning with the data having three teacher points in one-dimensional input $x$ and one-dimensional output $y$. We express these teacher signal points as $(\cdot)$, i.e., $(\hat{x}, \hat{y}) = (1, 4), (3, 5), (6, 2)$. This is a common situation in an adaptive antenna to adjust its weights by using pilot signals and an inverse problem to learn examples. If we expect appropriate outputs only at these teacher inputs, we do not need any learning process. Instead, we have only to make a table of (memorize) the input-output values shown in Fig. 1(b). However, what we actually expect is the input-output relationship shown by the curve(s) in Fig. 1(c), generating outputs even for inputs that were not used in the learning process. This is done by the generalization ability, and the curve shows the generalization characteristics. A good generalization characteristics are not obtained by simply nullifying the square error $\sum(y - \hat{y})^2$ where $y$ is the output value varying through the learning process. Moreover, we cannot say what curve is ideal before an actual task is given.

A small degree of freedom (DoF) often prevents the convergence of the error value to zero. Too large DoF, however, makes the curve wind, resulting in inappropriate estimation or prediction. Appropriate DoF is desired, which is realized by pruning or addition of network connections. We have also to check the meaningfulness of the alive connections by examining the independence among the signals going through the connections existing in parallel so that every connection works meaningfully without redundancy. In this sense, complex-valued connections and neurons hold a great advantage in dealing with waves and wave-related information. This is shown as follows.

B. Real $2 \times 2$ matrix representation

There are multiple manners to represent complex numbers. In 1835, Hamilton defined complex number $z$ in an algebraic manner as ordered pair of real numbers, $x$ and $y$, as $z \equiv (x, y)$ with the arithmetic of addition and multiplication as

$$(x_1, y_1) + (x_2, y_2) \equiv (x_1 + x_2, y_1 + y_2) \quad (1)$$

$$(x_1, y_1) \cdot (x_2, y_2) \equiv (x_1x_2 - y_1y_2, x_1y_2 + y_1x_2) \quad (2)$$

where the multiplication has a special characteristic. We can also use real $2 \times 2$ matrices to represent complex numbers [60], [61]. With every complex number $c = a + jib$, we associate the $C$-linear transformation

$$T_c : C \rightarrow C, \quad z \rightarrow cz = ax - by + j(bx + ay) \quad (3)$$

If we identify $C$ with $R^2$ by

$$z = (x + jy) = \begin{pmatrix} x \\ y \end{pmatrix} \quad (4)$$

it follows that

$$T_c \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (5)$$

That is, the linear transformation $T_c$ determined by $c = a + jib$ is represented by the matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Generally,
either. Then we can express a general input–output relationship as
\[
\begin{pmatrix}
  x_{1\text{OUT}} \\
  x_{2\text{OUT}}
\end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_{1\text{IN}} \\
  x_{2\text{IN}} \end{pmatrix} \tag{7}
\]
There exist a variety of possible mapping obtained through the learning because the number of parameters to be determined is larger than that of the conditions; i.e., the learning task is an ill-posed problem. The functional difference emerges as the difference in the generalization characteristics. For example, learning can result in a degenerate mapping shown in Fig. 3(b), which is often useless in practice.

In contract, let us consider the mapping learning task in the one-dimensional complex domain, which transforms a complex value \(x_{\text{IN}} = (x_{1\text{IN}}, x_{2\text{IN}})\) to another complex value \(x_{\text{OUT}} = (x_{1\text{OUT}}, x_{2\text{OUT}})\). Fig. 3(c) shows the complex-valued network, where the weight is a single complex value. The situation is expressed just like in (7) as
\[
\begin{pmatrix}
  x_{1\text{OUT}} \\
  x_{2\text{OUT}}
\end{pmatrix} = \begin{pmatrix} |w| \cos \theta & -|w| \sin \theta \\ |w| \sin \theta & |w| \cos \theta \end{pmatrix} \begin{pmatrix} x_{1\text{IN}} \\
  x_{2\text{IN}} \end{pmatrix} \tag{8}
\]
where \(\theta \equiv \arg(w)\). There are only two parameters \(|w|\) and \(\theta\). The learning result is given in Fig. 3(d). The DoF is lower, and the arbitrariness of solution is reduced, resulting in a simple combination of phase rotation and amplitude scaling. This result is interpreted also as the effectiveness of sparsity embedded in the number itself. This result is compatible with wave phenomenon.

The size of the above network is so small that the relationship between the parameter number and DoF is estimated very simply. In general, the dynamics of a neural network is determined by various parameters such as network structure, input–output data dimensions, and teacher signal numbers. However, the above characteristics of phase rotation and amplitude modulation are embedded in the complex-valued network as a universal elemental process of weighting.

It is true that the essential merit of neural networks in general lies in the high degree of freedom in learning and self-organization. However, if we know a priori that the objective quantities include "phase" and/or "amplitude," we can reduce possibly harmful portion of the freedom by employing a complex-valued neural network, resulting in a more meaningful generalization characteristic. The "rotation" in the complex multiplication works as an elemental process at the synapse, and it realizes the advantageous reduction of the degree of freedom. This is a type of sparsity built in the complex number itself.

IV. CHANNEL PREDICTION IN MOBILE COMMUNICATIONS: CVNN DEALS WITH THE COMPLEX CHANNEL AS AN ACTUAL ENTITY

We present an example of the improvement of neural dynamics by introducing physical picture into neural networks. The example is channel prediction in mobile communications. The details are given in Ref. [62].
Channel prediction is indispensable for future effective pre-equalization and/or transmission-power control in mobile communications to avoid harmful fading phenomenon in multipath environment. There have been many proposals in time domain and frequency domain including super-resolution methods. But they are not useful practically because of the poor prediction performance and/or the large calculation cost. Recently, however, we solved this problem by proposing a new method by combining the chirp z-transform (CZT) and a complex-valued neural network [62].

Fig. 4 is the schematic diagram showing our channel prediction. We transform the observation signal illustrated in Fig. 4(a) by using the CZT to obtain the spectrum with mitigation of discreteness as shown in (b). We realize the path separation, by marking the peaks in the magnitude, based on so-called Jakes model. We obtain the estimates of Doppler shift frequency $f_m$ and the amplitude $\hat{a}_m$ of path $m$, and also

Fig. 5 is the geometrical setup including two scatterers, a base station and a mobile user in the line of sight, which can cause severe fading [62].

Fig. 6 is the layered complex-valued neural network having a small size to reduce the calculation cost [62].
the phase $\hat{\theta}_m$ by referring to the phase spectrum shown in (c). Previously before the proposal, we predicted the channel based on the respective past parameters $f_m$, $\bar{a}_m$, $\bar{\phi}_m$ individually to obtain the present ($t = 0$) and future ($t \geq 0$) parameters $\hat{f}_m$, $\hat{a}_m$, $\hat{\phi}_m$ as

$$ \hat{f}_m, \hat{a}_m, \hat{\phi}_m \overset{\text{Predict}}{\leftarrow} f_m, a_m, \phi_m $$

(9)

$$ \hat{c}(t) = \sum_m \hat{a}_m e^{j2\pi f_m t + \phi_m} $$

(10)

It worked well mostly. However, it sometimes failed when the situation is very difficult. An example is shown in Fig. 5 where the receiver runs just beside two scatters located near to each other.

In our newly proposed method, we consider the channel as a unified complex-valued entity, and treated it by using the complex-valued layered neural network shown in Fig. 6 in combination with CZT generating a series of past channel states $\hat{c}_m(t - 1), ..., \hat{c}_m(t - I_{\text{ML}})$ as

$$ \hat{c}_m(t) \equiv \hat{a}_m(t)e^{j\hat{\phi}_m(t)} $$

$$ (\hat{\theta}_m(t) \equiv 2\pi f_m(t) + \hat{\phi}_m(t) ) $$

(11)

$$ \hat{c}_m(t) \overset{\text{Predict}}{\leftarrow} \hat{c}_m(t - 1), ..., \hat{c}_m(t - I_{\text{ML}}) $$

(12)

We slide the prediction window by repeating this prediction to come to the future time slot. Finally we unite the respective path states $\hat{c}_m(t)$ to predict the channel state in the future as

$$ \hat{c}(t) = \sum_m \hat{c}_m(t) $$

(13)

Prediction based on the estimated path states $\hat{c}_m(t)$ appeared more difficult than the prediction on the respective parameters $f_m$, $\bar{a}_m$, $\bar{\phi}_m$. However, our physical experiments showed a channel state with near-circular time evolution, which matches the representation with the sparsity built in the complex number itself. This fact suggests that the complex-amplitude prediction is better than the prediction based on respective Doppler frequency, magnitude and phase. Fig. 7 shows the bit error rate (BER) obtained for the difficult case shown in Fig. 5 when we employ various prediction methods, namely, time-domain linear prediction, time-domain real-time recurrent-learning CVNN-based prediction, CZT only without prediction process, CZT frequency-domain linear prediction, CZT frequency-domain Lagrange-based prediction, CZT frequency-domain auto-regression-model-based prediction, CZT frequency-domain real-time recurrent-learning CVNN-based prediction (the proposal), CZT frequency-domain multilayer-CVNN-based prediction, and ideal case where we assume that we could obtain actual channel state (impossible in reality), respectively.

It is shown that our proposed method that combined the CZT and the multilayer CVNN realizes the lowest BER. The calculation cost is small because of its successive online learning. The result directly presents the high performance of the CVNN when it deals with wave information.

V. SUMMARY

This paper discussed the origin of the advantages of CVNNs. We focused on the generalization ability originating from the sparsity existing in the complex number itself. The merit is clearly presented in the performance comparison in channel prediction for mobile communications.

REFERENCES


