CHAPTER 1

APPLICATION FIELDS AND FUNDAMENTAL MERITS OF COMPLEX-VALUED NEURAL NETWORKS

AKIRA HIROSE

The University of Tokyo, Tokyo, Japan

This chapter presents historical and latest advances in applications of complex-valued neural networks (CVNNs) first. Then it also shows one of the most important merits of CVNNs, namely, the suitability for adaptive processing of coherent signals.

1.1 INTRODUCTION

This chapter presents historical and latest advances in applications of complex-valued neural networks (CVNNs) first. Then it also shows one of the most important merits of CVNNs, namely, the suitability for adaptive processing of coherent signals.

CVNNs are effective and powerful in particular to deal with wave phenomena such as electromagnetic and sonic waves, as well as to process wave-related information. Regarding the history of CVNNs, we can trace back to the middle of the 20th century. The first introduction of phase information in computation was made by Eiichi Goto in 1954 in his invention of "Parametron" [17, 18, 61]. He utilized the phase of a high-frequency carrier to represent binary or multivalued informa-

Complex-Valued Neural Networks: Advances and Applications. Edited by Akira Hirose

tion. However, the computational principle employed there was "logic" of Turing type, or von Neumann type, based on symbol processing, so that he could not make further extensive use of the phase. In the present CVNN researches, contrarily, the researchers extend the world of computation to pattern processing fields based on a novel use of the structure of complex-amplitude (phase and amplitude) information.

We notice that the above feature is significantly important when we give thought to the fact that various modern technologies centered on electronics orient toward coherent systems and devices rather than something incoherent. The feature will lead to future general probability statistics, stochastic methods, and statistical learning and self-organization framework in coherent signal processing and information analysis. The fundamental idea is applicable also to hypercomplex processing based on quaternion, octonion, and Clifford algebraic networks.

Some parts of the following contents of this chapter were published in detail in *the Journal of Society of Instrument and Control Engineers* [29], *the Frontiers in Electrical and Electronic Engineering in China* [28], and *IEEE Transactions in Neural Networks and Learning Systems* [35].

1.2 APPLICATIONS OF COMPLEX-VALUED NEURAL NETWORKS

Complex-valued neural networks (CVNNs) have become widely used in various fields. The basic ideas and fundamental principles have been published in several books in recent years [27, 22, 26, 41, 53, 2]. The following subsections present major application fields.

1.2.1 Antenna Design

The most notable feature of CVNNs is the compatibility with wave phenomena and wave information related to, for example, electromagnetic wave, lightwave, electron wave, and sonic wave [28]. Application fields include adaptive design of antennas such as patch antennas for microwave and millimeter wave. Many researches have been reported on how to determine patch-antenna shape and sub-element arrangement, as well as on the switching patterns of the sub-elements [46, 10, 47]. A designer assigns desired frequency-domain characteristics of complex amplitude, or simply amplitude, such as transmission characteristics, return loss, and radiation patterns. A CVNN mostly realizes a more suitable design than a real-valued network does even when he/she presents only simple amplitude. The reason lies in the elemental dynamics consisting of phase rotation (or time delay $\times$ carrier frequency) and amplitude increase or decrease, based on which dynamics the CVNN learning or self-organization works. As a result, the generalization characteristics (error magnitude at nonlearning points in supervised learning) and the classification manner often become quite different from those of real-valued neural networks [28, 35]. The feature plays the most important role also in other applications referred to below.

1.2.2 Estimation of Direction of Arrival and Beamforming

The estimation of direction of arrival (DoA) of electromagnetic wave using CVNNs has also been investigated for decades [67, 6]. A similar application field is the beamforming. When a signal has a narrow band, we can simply employ Huygens' principle. However, in an ultra-wideband (UWB) system, where the wavelength is distributed over a wide range, we cannot assume a single wavelength, resulting in unavailability of Huygens' principle. To overcome this difficulty, an adaptive method based on a CVNN has been proposed [60] where a unit module consists of a tapped-delay-line (TDL) network.

1.2.3 Radar Imaging

CVNNs are widely applied in coherent electromagnetic-wave signal processing. An area is adaptive processing of interferometric synthetic aperture radar (InSAR) images captured by satellite or airplane to observe land surface [59, 65]. There they aim at solving one of the most serious problems in InSAR imaging that there exist many rotational points (singular points) in the observed data so that the height cannot be determined in a straightforward way.

Ground penetrating radar (GPR) is another field [21, 66, 43, 44, 49, 34]. GPR systems usually suffer from serious clutter (scattering and reflection from non-target objects). Land surface as well as stones and clods generate such heavy clutter that we cannot observe what are underground if we pay attention only to the intensity. Complex-amplitude texture often provides us with highly informative features that can be processed adaptively in such a manner that we do in our early vision.

1.2.4 Acoustic Signal Processing and Ultrasonic Imaging

Another important application field is sonic and ultrasonic processing. Pioneering works were done into various directions [69, 58]. The problem of singular points exists also in ultrasonic imaging. They appear as speckles. A technique similar to that used in InSAR imaging was successfully applied to ultrasonic imaging [51].

1.2.5 Communications Signal Processing

In communication systems, we can regard CVNNs as an extension of adaptive complex filters, i.e., modular multiple-stage and nonlinear version. From this viewpoint, several groups work hard on time-sequential signal processing [15, 16], blind separation [68], channel prediction [12], equalization [63, 36, 55, 40, 33, 7, 8], and channel separation in multiple-input multiple-output (MIMO) systems [37]. Relevant circuit realization [13] is highly inspiring not only as working hardware but also for understanding of neural dynamics.

1.2.6 Image Processing

There are many ideas based on CVNNs also in image processing. An example is the adaptive processing for blur compensation by identifying point scattering function in the frequency domain [3]. In such a frequency-domain processing of images, we often utilize the fact that the phase information in frequency domain corresponds to position information in spatial domain. On the other hand, CVNN spatial-domain processing is also unique and powerful. A highly practical proposal was made for quick gesture recognition in smart phones by dealing with finger angle information adaptively by a CVNN [19]. Biological imaging is another expanding field. There we can find, for example, a classification of gene-expression stages in gene images [1], along with adaptive segmentation of magnetic resonance image (MRI) by placing a dynamic boundary curve (so-called "snake") in the obtained complex-amplitude MRI image for segmentation of blood vessels and other organs [20]. Since there are various types of active and coherent imaging systems in medicine, we can expect further applications of CVNNs to deal with complex-amplitude images.

1.2.7 Social Systems Such as Traffic and Power Systems

Recent applications expand more multi-directionally even to social systems. In traffic systems, a CVNN will be effectively used for controlling mutual switching timing of traffic lights in complicatedly connected driving roads [50]. Since traffic lights have periodic operation, some CVNN dynamics is suitable for their adaptive control. Green energy and smart grid are also the fields. A CVNN-based prediction of wind strength and direction has been demonstrated for efficient electric power generation [14] in which amplitude and phase in the complex plane represent the strength and the direction, respectively.

1.2.8 Quantum Devices Such as Superconductive Devices

Applications to quantum computation using quantum devices such as superconductivity have also been investigated in many groups [57, 39, 48]. Their results suggest the future realization of intrinsically non-von Neumann computers including pattern-information representing devices. Conventional quantum computation is strictly limited in its treatable problems. Contrarily, CVNN-based quantum computation can deal with more general problems, which leads to wider applications of quantum computation.

1.2.9 Optical/Lightwave Information Processing Including Carrier-Frequency Multiplexing

Learning optical and lightwave computer is another field of CVNN applications. There are researches such as frequency-domain multiplexed learning [38] and real-time generation of a three-dimensional holographic movie for interactively controllable optical tweezers [32, 62]. In these networks, a signal has its carrier frequency,

equivalent to a band signal in communications, and therefore the learning and processing dynamics is controllable by modulating the carrier frequency. The idea can be adapted to complex filters. It led to a novel developmental learning of motion control combined with reinforcement learning [30]. The success suggests further a possible influence of frequency modulation of brain wave on biological brain activity, indicating a new door to CVNN-related physiology.

1.2.10 Hypercomplex-Valued Neural Networks

Hypercomplex-valued neural networks have also been actively investigated [5]. An example is the adaptive learning in three-dimensional color space by using quaternion [45]. An adaptive super-high-sensitive color camera (so-called night vision) has been produced that realizes a compensation of nonlinear human color-vision characteristics in extremely dark environment. More generalized hypercomplex networks, namely, Clifford algebraic neural networks, are also discussed very actively in, e.g., special sessions in conferences [54].

1.3 WHAT IS A COMPLEX NUMBER?

In this section, we look back the history of complex numbers to extract the essence influential in neural dynamics.

1.3.1 Geometric and Intuitive Definition

Throughout history, the definition of the complex number has changed gradually [11]. In the 16th century, Cardano tried to work with imaginary roots in dealing with quadratic equations. Afterward, Euler used complex numbers in his calculations intuitively and correctly. It is said that by 1728 he knew the transcendental relationship $i \log i = -\pi/2$. The Euler formulae appear in his book as

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \quad \text{and} \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i} \tag{1.1}$$

In 1798, Wessel described representation of the points of a plane by complex numbers to deal with directed line segments. Argand also interpreted $\sqrt{-1}$ as a rotation through a right angle in the plane, and he justified this idea on the ground that two $\sqrt{-1}$ rotations yields a reflection, i.e., -1. It is also believed that, in early 1749, Euler already had a visual concept of complex numbers as points of a plane. He described a number x on a unit circle as $x = \cos g + i \sin g$, where g is an arc of the circle. Gauss was in full possession of the geometrical theory by 1815. He proposed to refer to $+1$, -1, and $\sqrt{-1}$ as direct, inverse, and lateral unity, instead of positive, negative, and imaginary or "impossible" elements.

1.3.2 Definition as Ordered Pair of Real Numbers

The geometrical representation is intuitively simple and visually understandable, but may be weak in strictness. In 1835, Hamilton presented the formal definition of the complex number as an "ordered pair of real numbers," which also led to the discovery of quaternions, in his article entitled "Theory of conjugate functions, or algebra as the science of pure time." He defined addition and multiplication in such a manner that the distributive, associative, and commutative laws hold. The definition as the ordered pair of real numbers is algebraic, and it can be stricter than the intuitive rotation interpretation.

At the same time, the fact that a complex number is defined by two real numbers may lead present-day neural-network researchers to consider a complex network equivalent to just a doubled-dimension real-number network effectively. However, in this paper, the authors would like to clarify the merit by focusing on the rotational function even with this definition.

Based on the definition of the complex number as an ordered pair of real numbers, we represent a complex number z as

$$z \equiv (x, y) \tag{1.2}$$

where x and y are real numbers. Then the addition and multiplication of z_1 and z_2 are defined in *complex domain* as

$$(x_1, y_1) + (x_2, y_2) \equiv (x_1 + x_2, y_1 + y_2) \tag{1.3}$$

$$(x_1, y_1) \cdot (x_2, y_2) \equiv (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2) \tag{1.4}$$

As a reference, the addition and multiplication (as a step in correlation calculation, for example) of *two-dimensional real values* is expressed as

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \tag{1.5}$$

$$(x_1, y_1) \cdot (x_2, y_2) = (x_1 x_2, y_1 y_2) \tag{1.6}$$

In the comparison, the addition process is identical. Contrarily, the complex multiplication seems quite artificial, but this definition (1.4) brings the complex number with its unique function, that is, the angle rotation, as well as amplitude amplification/attenuation, which are the result of the intermixture of the real and imaginary components.

It is easily verified that the commutative, associative, and distributive laws hold. We have the unit element $(1, 0)$ and the inverse of z $(\neq 0)$, which is

$$\begin{aligned} z^{-1} &\equiv \left(\frac{x}{x^2 + y^2}, \frac{-y}{x^2 + y^2} \right) \\ &= \left(\frac{x}{|z|^2}, \frac{-y}{|z|^2} \right) \end{aligned} \tag{1.7}$$

where $|z| \equiv \sqrt{x^2 + y^2}$.

1.3.3 Real 2×2 Matrix Representation

We can also use real 2×2 matrices, instead of the ordered pairs of real numbers, to represent complex numbers [11, 9]. With every complex number $c = a + ib$, we associate the C-linear transformation

$$T_c : C \to C, \quad z \mapsto cz = ax - by + i(bx + ay) \tag{1.8}$$

which includes a special case of $z \to iz$ that maps 1 into i, i into -1, ..., with a rotation with right angle each. In this sense, this definition is a more precise and general version of Argand's interpretation of complex numbers. If we identify C with R^2 by

$$z = x + iy = \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.9}$$

it follows that

$$\begin{aligned} T_c \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix} \\ &= \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \end{aligned} \tag{1.10}$$

In other words, the linear transformation T_c determined by $c = a + ib$ is described by the matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Generally, a mapping represented by a 2×2 matrix is noncommutative. However, in the present case, it becomes *commutative*. By this real matrix representation, the imaginary unit i in C is given as

$$I \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad I^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -E \tag{1.11}$$

In the days of Hamilton, we did not have matrices yet. Even after the advent of matrices, it is very rare to define complex numbers in terms of real 2×2 matrices [11] (Chapter 3, §2, 5.), [9] . The introduction of complex numbers through 2×2 matrices has the advantage, over introducing them through ordered pairs of real numbers, that it is unnecessary to define an ad hoc multiplication. What is most important is that this matrix representation clearly expresses the function specific to the complex numbers—that is, the rotation and amplification or attenuation as

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{1.12}$$

where r and θ denote amplification/attenuation of amplitude and rotation angle applied to signals, respectively, in the multiplication calculation. On the other hand, addition is rather plain. The complex addition function is identical to that in the case of doubled-dimension real numbers.

In summary, the phase rotation and amplitude amplification/attenuation are the most important features of complex numbers.

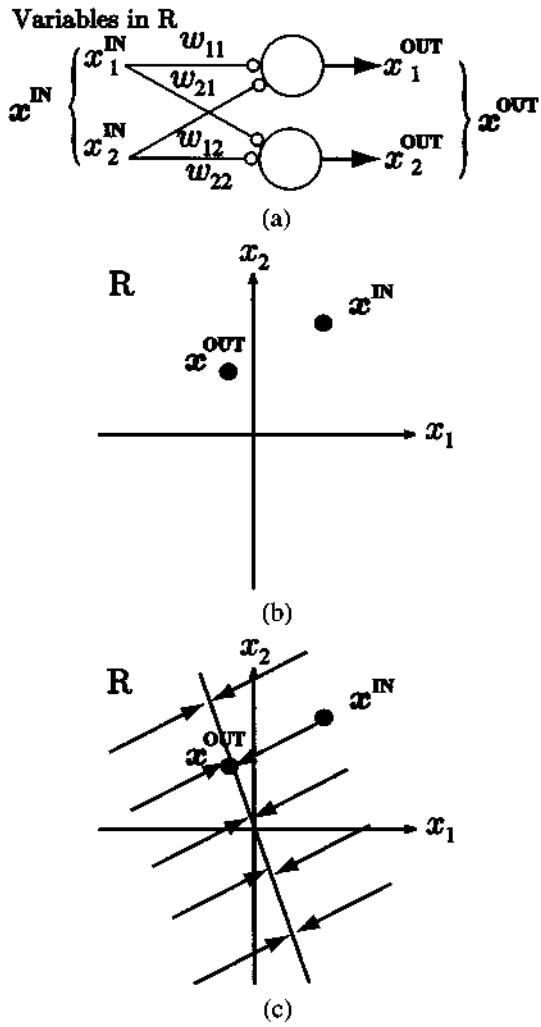


Figure 1.1 (a) A simple real-valued single-layered two-input two-output feedforward network to learn (b) a mapping that maps x^{IN} to x^{OUT} and (c) a possible but degenerate solution that is often unuseful [28].

1.4 COMPLEX NUMBERS IN FEEDFORWARD NEURAL NETWORKS

We consider intuitively what feature emerges in the dynamics of complex-valued neural networks. Here we first take a layered feedforward neural network. Then we consider metrics in correlation learning.

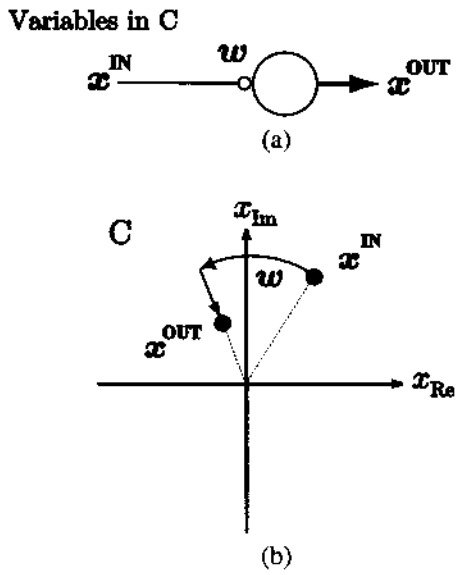


Figure 1.2 (a) A complex-valued neural network seemingly identical to Fig. 1.1(a) to learn the mapping shown in Fig. 1.1(b), and (b) a solution obtained in this small-degree-of-freedom case [28].

1.4.1 Synapse and Network Function in Layered Feedforward Neural Networks

In wave-related adaptive processing, we often obtain excellent performance with learning or self-organization based on the CVNNs. As already mentioned, the reason depends on situations. However, the discussion in Section 1.3 suggests that the origin lies in the complex rule of arithmetic. That is to say, the merit arises from the functions of the four fundamental rules of arithmetic of complex numbers, in particular the multiplication, rather than the representation of the complex numbers, which can be geometric, algebraic, or in matrices. Moreover, the essence of the complex numbers also lies in the characteristic multiplication function, the phase rotation, as overviewed in Section 1.3 [27].

Let us consider a very simple case shown in Fig. 1.1(a), where we have a single-layer 2-input 2-output feedforward neural network in real number. For simplicity, we omit the possible nonlinearity at the neurons, i.e., the activation function is the identity function, where the neurons have no threshold. We assume that the network should realize a mapping that transforms an input $\boldsymbol{x}^{IN}$ to an output $\boldsymbol{x}^{OUT}$ in Fig. 1.1(b) through supervised learning that adjusts the synaptic weights w_{ji}. Simply, we have only a single teacher pair of input and output signals. Then we can

describe a general input–output relationship as

$$\begin{pmatrix} x_1^{OUT} \\ x_2^{OUT} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1^{IN} \\ x_2^{IN} \end{pmatrix} \tag{1.13}$$

We have a variety of possible mapping obtained by the learning because the number of parameters to be determined is larger than the condition; 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. 1.1(c), which is often unuseful in practice.

Next, let us consider the mapping learning task in the one-dimensional complex domain, which transforms a complex value $\boldsymbol{x}^{IN} = (x_1^{IN}, x_2^{IN})$ to another complex value $\boldsymbol{x}^{OUT} = (x_1^{OUT}, x_2^{OUT})$. Figure 1.2(a) shows the complex-valued network, where the weight is a single complex value. The situation is expressed just like in (1.13) as

$$\begin{pmatrix} x_1^{OUT} \\ x_2^{OUT} \end{pmatrix} = \begin{pmatrix} |w|\cos\theta & -|w|\sin\theta \\ |w|\sin\theta & |w|\cos\theta \end{pmatrix} \begin{pmatrix} x_1^{IN} \\ x_2^{IN} \end{pmatrix} \tag{1.14}$$

where $\theta \equiv \arg(w)$. The degree of freedom is reduced, and the arbitrariness of the solution is also reduced. Figure 1.2(b) illustrates the result of the learning. The mapping is a combination of phase rotation and amplitude attenuation. This example is truly an extreme. 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.

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 characteristics. 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 feature corresponds not only to the geometrical intuitive definition of complex numbers but also to the Hamilton's definition by ordered pairs of real numbers, or the real 2×2 matrix representation.

Though we considered a small feedforward network in this section, the conclusion is applicable also to other CVNNs such as complex-valued Hebbian-rule-based network and complex correlation learning networks, where the weight is updated by the multiplication results. The elemental process of phase rotation and amplitude modulation results in the network behavior consistent with phase rotation and amplitude modulation in total. The nature is a great advantage when we deal with not only waves such as electromagnetic wave and lightwave, but also arbitrary signals with the Fourier synthesis principle, or in the frequency domain through the Fourier transform.

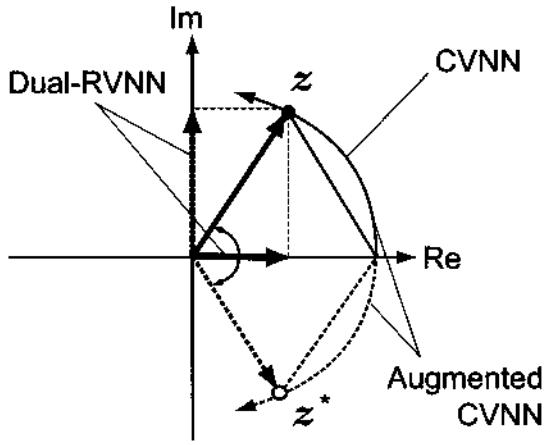


Figure 1.3 Conceptual illustration of the relationship among bases in the respective neural networks to deal with complex signal z [35].

1.4.2 Circularity

The circularity of the signals to be processed is also an important factor. To deepen the discussion, we refer to the wide sense linear (or widely linear: WL) systems which introduce conjugate signals in addition to direct complex signals [56, 41]. WL systems well learn complex data distributed anisotropically in the complex plane, i.e., noncircular data. For example, it is useful for predicting wind strength and direction by assuming the axes of the complex number plane represent north, south, east, and west, and the distance from the origin expresses the strength. Augmented complex-valued neural networks have been proposed in such a context [64]. Wind has high anisotropy in general. The augmented complex-valued networks does not lead to the reduction of the degree of freedom. The degree is the same as that of real-valued networks, resulting in dynamics similar to that of real-valued ones [42].

Figure 1.3 is a conceptual illustration showing the bases of the respective networks. The number of the bases of the augmented complex networks becomes that of the real-valued networks back, and its dynamics approaches that of real networks. This situation is analogous to the fact that the combination of positive and negative frequency spectra generates almost real-valued signals. In other words, if we compare the relationship to the polarization of lightwave, we come to the following. Complex-valued neural networks deal with only right- or left-handed circular polarized light, which are suitable for circular signal processing. Note that the signal in total can be out of complete circularity, but only each frequency component has the circularity. Since any waveform can be synthesized by sinusoidal components through Fourier synthesis, the signals that the complex networks can deal with are not limited to completely coherent signals. In contrast, the augmented complex-valued networks deal with both the right- and left-handed circular polarized light.

They are more flexible because of the larger degree of freedom, which is too much for circular signals. Dual univariate networks have the same degree of freedom; however, in this case, the bases are linear polarization corresponding to the real and imaginary parts, instead of the right- and left-handed circular bases in the augmented networks. In this manner, they are similar to each other.

Consequently, complex-valued neural networks are suitable for processing analytic signals, which consist of a real component and its consistent imaginary component that has the same amplitude but 90-degree shifted phase. The analytic signal is essentially circular. Analytic signals exist widely in electronics—for example, at the output of heterodyne or homodyne mixers and at the output of digital signal processing using the Hilbert transform. Complex-valued networks have the ability to process such analytic signals appropriately.

1.5 METRIC IN COMPLEX DOMAIN

1.5.1 Metric in Complex-Valued Self-Organizing Map

Among various neurodynamics in the complex domain, the complex-valued self-organizing maps (CSOMs) may possess fewer features which reflect the complex multiplication mentioned in Section 1.4 since most of SOMs have two subprocesses in the operation, i.e., winner determination and weight update, both of which may consist of only addition and subtraction in its arithmetic without any multiplication that utilizes the complex nature of phase rotation.

However, the circumstances depend on the metric we use to determine the dynamics. If we employ complex inner product, instead of conventional Euclidean metric in double-dimensional real space, we can utilize the characteristics specific to complex space [4]. The general dynamics of a SOM will be explained in Section 1.5. In this section, we discuss the metric we use in feature vector space.

1.5.2 Euclidean Metric

In SOM in general, the metric most widely used to determine the *winner* neuron whose weight $\boldsymbol{w}_c$ is nearest to an input feature vector $\boldsymbol{z}$ is the Euclidean metric. Even in a complex-valued SOM (CSOM) where $\boldsymbol{z}$ and $\boldsymbol{w}$ are complex, we can express them with imaginary unit i as

$$\boldsymbol{z} \equiv \begin{bmatrix} |z_1| \exp(i\theta_1) \\ |z_2| \exp(i\theta_2) \\ \vdots \end{bmatrix} \tag{1.15}$$

$$\boldsymbol{w}_c \equiv \begin{bmatrix} |w_{c\,1}| \exp(i\psi_{c\,1}) \\ |w_{c\,2}| \exp(i\psi_{c\,2}) \\ \vdots \end{bmatrix} \tag{1.16}$$

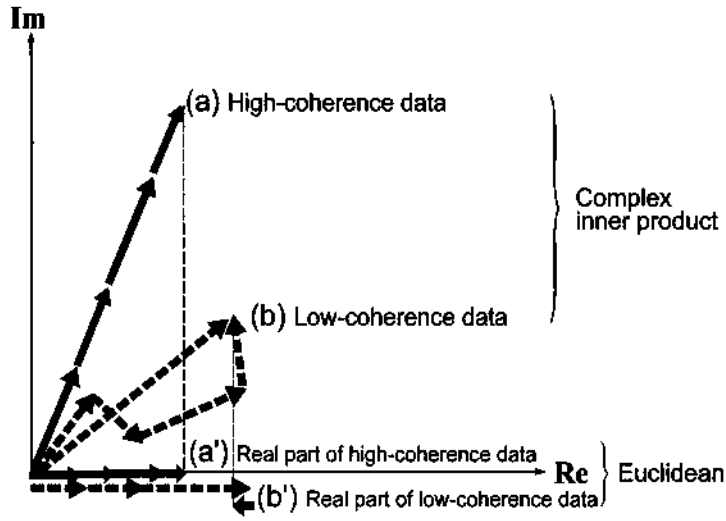


Figure 1.4 Conceptual illustration to compare the inner product $\boldsymbol{z}^*\boldsymbol{w}_c$ and the real-part inner product $\mathbf{Re}(\boldsymbol{z}^*\boldsymbol{w}_c)$ to calculate $||\boldsymbol{z}-\boldsymbol{w}_c||^2$ [4].

The Euclidean process to choose a winner is expressed as

$$\hat{c} = \arg\min_c ||\boldsymbol{z}-\boldsymbol{w}_c|| \quad (c : \text{class index}) \tag{1.17}$$

where $\arg\min_c \cdots$ chooses a c that minimizes $\cdots$, and $||\cdot||$ denotes norm (amplitude), i.e.,

$$\begin{aligned} ||\boldsymbol{z}-\boldsymbol{w}_c||^2 &= (\boldsymbol{z}-\boldsymbol{w}_c)^* (\boldsymbol{z}-\boldsymbol{w}_c) \\ &= ||\boldsymbol{z}||^2 + ||\boldsymbol{w}_c||^2 - (\boldsymbol{z}^*\boldsymbol{w}_c + \boldsymbol{w}_c^*\boldsymbol{z}) \\ &= ||\boldsymbol{z}||^2 + ||\boldsymbol{w}_c||^2 - 2\mathbf{Re}(\boldsymbol{z}^*\boldsymbol{w}_c) \end{aligned} \tag{1.18}$$

Though (1.18) deals with complex numbers, this arithmetic is identical to the calculation of real-valued Euclidean distance and also of the real-valued inner product; i.e., when $\boldsymbol{x}, \boldsymbol{w}_c \in \mathbf{R}^m$,

$$\begin{aligned} ||\boldsymbol{x}-\boldsymbol{w}_c||^2 &= ||\boldsymbol{x}||^2 + ||\boldsymbol{w}_c||^2 - 2\boldsymbol{x}^T\boldsymbol{w}_c \\ &= ||\boldsymbol{x}||^2 + ||\boldsymbol{w}_c||^2 - 2\sum_i |x_i||w_{c\,i}|\cos(\psi_{c\,i} - \theta_i) \end{aligned} \tag{1.19}$$

Then, when $||\boldsymbol{z}||^2$ and $||\boldsymbol{w}_c||^2$ are almost constants, as is often the case, and/or we pay attention to phase information, the distance (1.19) is detemined by the cosine component (real part) $|z_i||w_{c\,i}|\cos(\psi_{c\,i} - \theta_i)$.

1.5.3 Complex Inner-Product Metric

Instead, we can also employ a complex inner-product metric for use in determination of a winner in the CSOM as

$$\hat{c} = \arg\max_{c} \left(\left| \frac{\boldsymbol{z}^* \, \boldsymbol{w}_c}{||\boldsymbol{z}|| ||\boldsymbol{w}_c||} \right| \right) \quad (c : \text{class index}) \tag{1.20}$$

This process is better understandable in equations by employing the polar representation. That is, the numerator of the complex-valued inner product (1.20) is given as

$$\boldsymbol{z}^* \, \boldsymbol{w}_c = \sum_i \left(|z_i| \exp(-i\theta_i) \right) \left(|w_{c\,i}| \exp(i\psi_{c\,i}) \right) \tag{1.21}$$

$$= \sum_i |z_i| |w_{c\,i}| \exp(i(\psi_{c\,i} - \theta_i)) \tag{1.22}$$

where the summation takes into account the phase values directly, that is, the direction of the arrows [4].

In other words, the metric (1.22) takes both the cosine and sine components (real and imaginary parts) into consideration. That is, when we express the vectors as $\boldsymbol{z} = (x_1 + iy_1, x_2 + iy_2, ...)$ and $\boldsymbol{w} = (u_1 + iv_1, u_2 + iv_2, ...)$, omitting suffix c, we obtain

$$\begin{aligned} \boldsymbol{z}^* \, \boldsymbol{w} &= [x_1 - iy_1 \;\; x_2 - iy_2 \;\; ... \;] \, [u_1 + iv_1 \;\; u_2 + iv_2 \;\; ... \;]^T \\ &= x_1u_1 + y_1v_1 + x_2u_2 + y_2v_2 + ... \qquad \Longleftarrow \text{cos component} \\ &\quad + i\,(x_1v_1 - y_1u_1 + x_2v_2 - y_2u_2 + ...) \quad \Longleftarrow \text{sin component} \end{aligned} \tag{1.23}$$

1.5.4 Comparison Between Complex Inner Product and Euclidean Distance

Figure 1.4 is a conceptual illustration to show the merit of this complex inner-product metric. In active imaging such as the ground penetrating radars described in Section 1.2.3, we obtain coherent signals consisting of amplitude and phase. The feature vector is defined in complex domain. For a set of high-coherence signals, i.e., signals having similar phases, the summation to generate inner product grows straightforward as shown by arrows (a) in Fig. 1.4. Contrarily, in a low-coherence case, having random phases, the summation does not grow so much as shown by arrows (b). This effect emerges also in the Euclidean metric to some extent. However, the Euclidean metric is related only to the cosine component as shown in Fig. 1.4(a′) and (b′), resulting in a partial treatment of phase directions. The evaluation results can be different from (a) and (b). The complex inner-product metric is then more sensitive to signal coherence and, therefore, enhances the distinction among various objects compared with the case of Euclidean metric described below.

In addition, the complex inner product is inherently less sensitive to the norm of signal vectors. This is simply because of the normalization. It is desirable in partic-

ular in coherent imaging systems where we often suffer from distortion in intensity caused by the mirror glaring and speckles.

1.5.5 Metric in Correlation Learning

Correlation learning, used widely in neural networks such as associative memories [24] described in Section 1.5.5, also possess the same feature of the complex-valued learning. The correlation learning embeds the correlation between output signals $\boldsymbol{z}_s$ and input signals $\boldsymbol{z}_t$ in synaptic weights $\boldsymbol{w}$. For simplicity of expression, we consider one of the output signals z_s out of $\boldsymbol{z}_s$. As shown in detail in Section 1.5.5, the learning dynamics is expressed as

$$\tau \frac{d\boldsymbol{w}}{dt} = -\boldsymbol{w} + z_s\, \boldsymbol{z}_t^* \tag{1.24}$$

where τ is the learning time constant in the time t domain. Various pairs of input $\boldsymbol{z}_t$ and output z_s teacher signals are presented to the network for the training. The correlation is accumulated into $\boldsymbol{w}$, converging at

$$\boldsymbol{w} \longrightarrow K < z_s\, \boldsymbol{z}_t^* > \tag{1.25}$$

where K is a real constant.

Here we express the teacher signal pairs in real and imaginary parts as

$$z_s = x_s + jy_s \tag{1.26}$$

$$\boldsymbol{z}_t = [x_{t1} + jy_{t1}, x_{t2} + jy_{t2}, \ldots, x_{tN} + jy_{tN}]^T \tag{1.27}$$

where j and N are the imaginary unit and the input terminal number. Then the product in the correlation in (1.25) is rewritten as

$$\begin{aligned} z_s\, \boldsymbol{z}_t^* = \; & [\,(x_s x_{t1} + y_s y_{t1}) + j(y_s x_{t1} - x_s y_{t1}), \\ & (x_s x_{t2} + y_s y_{t2}) + j(y_s x_{t2} - x_s y_{t2}), \\ & \qquad \ldots \\ & (x_s x_{tN} + y_s y_{tN}) + j(y_s x_{tN} - x_s y_{tN})\,]^T \end{aligned} \tag{1.28}$$

The real and imaginary parts mix with each other. The meaning becomes obvious when we express the pixel values in amplitude and phase as

$$z_s = r_s e^{i\theta_s} \tag{1.29}$$

$$\boldsymbol{z}_t = [r_{t1} e^{j\theta_{t1}}, r_{t2} e^{j\theta_{t2}}, \ldots, r_{tN} e^{j\theta_{tN}}]^T \tag{1.30}$$

and rewrite (1.28) as

$$z_s\, \boldsymbol{z}_t^* = [\, r_s r_{t1} e^{j(\theta_s - \theta_{t1})}, r_s r_{t2} e^{j(\theta_s - \theta_{t2})}, \ldots, r_s r_{tN} e^{j(\theta_s - \theta_{tN})}]^T \tag{1.31}$$

The product yields the phase difference as well as the amplitude product, which is compatible with the signal circularity.

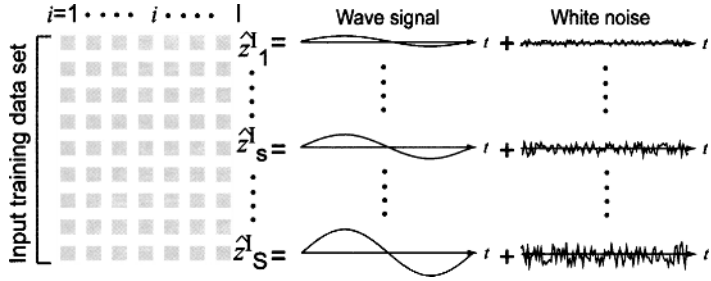


Figure 1.5 A set of teacher signals [35] (See color insert.).

On the contrary, if we regard the neural network as a real-valued network having double input terminals and two output neurons corresponding to real and imaginary parts, the dynamics for double-dimensional real signals z_s and z_t are expressed as

$$z_s = [x_s, y_s] \tag{1.32}$$

$$z_t = [x_{t1}, y_{t1}, x_{t2}, y_{t2}, \cdots, x_{tN}, y_{tN}]^T \tag{1.33}$$

and the product as a step to calculate correlation becomes

$$z_s z_t = [\, x_s x_{t1},\ y_s y_{t1},\ x_s x_{t2},\ y_s y_{t2},\ \cdots,\ x_s x_{tN},\ y_s y_{tN}]^T \tag{1.34}$$

We can find that the product (1.34) is different from (1.28) or (1.31). That is, the dynamics of the real-valued network is completely different from that of the complex-valued one. The difference originates from the very basic arithmetic operation, and is therefore very fundamental. This property may also be called circularity as one of the characteristics of the complex-valued neural network. The circularity is one of the most essential features of the complex-valued neural networks.

1.6 EXPERIMENTS TO ELUCIDATE THE GENERALIZATION CHARACTERISTICS

To elucidate the generalization characteristics in feedforward layered neural networks described in Section 1.4, we conducted a set of experiments. The details were reported in Ref. 35. The outline is explained as follows.

- Input signals: Weighted summation of the following (A) and (B) as shown in Fig. 1.5.
  (A) Sinusoid: completely coherent signal.
  (B) White Gaussian noise (WGN): completely incoherent data having random amplitude and phase (or real and imaginary parts).

- Task to learn: Identity mapping, which is expected to show the learning characteristics most clearly for the above signals with various degrees of coherence.

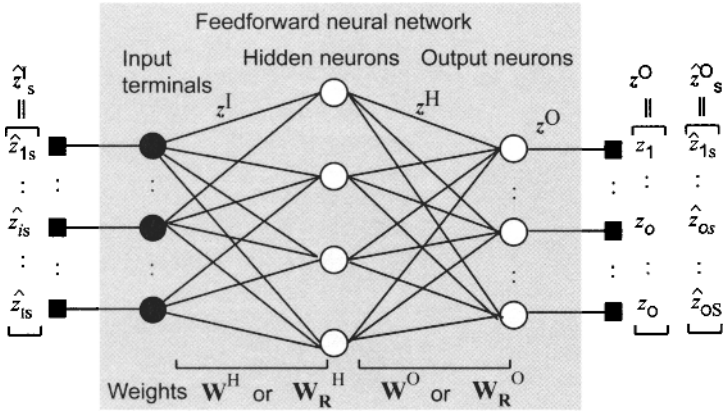


Figure 1.6 Basic construction of the complex- and real-valued feedforward neural networks [35]. (See color insert.)

- Evaluation of generalization: Observation of the generalization error when the input signals shift in time, or when the amplitude is changed.

1.6.1 Forward Processing and Learning Dynamics

1.6.1.1 Complex-Valued Neural Network (CVNN) Figure 1.6 shows the general construction of the neural network to be considered here. It is a layered feedforward network having input terminals, hidden neurons, and output neurons. In a CVNN, we first employ a phase-amplitude-type sigmoid activation function and the teacher-signal-backpropagation learning process, [23, 31] with notations of

$$z^{\mathrm{I}} = [z_1, ..., z_i, ..., z_I, z_{I+1}]^T \quad \text{(Input signal vector)} \tag{1.35}$$

$$z^{\mathrm{H}} = [z_1, ..., z_h, ..., z_H, z_{H+1}]^T \quad \text{(Hidden-layer output signal vector)} \tag{1.36}$$

$$z^{\mathrm{O}} = [z_1, ..., z_o, ..., z_O]^T \quad \text{(Output-layer signal vector)} \tag{1.37}$$

$$\mathbf{W}^{\mathrm{H}} = [w_{hi}] \quad \text{(Hidden neuron weight matrix)} \tag{1.38}$$

$$\mathbf{W}^{\mathrm{O}} = [w_{oh}] \quad \text{(Output neuron weight matrix)} \tag{1.39}$$

where $[\cdot]^T$ means transpose. In (1.38) and (1.39), the weight matrices include additional weights $w_{h\,I+1}$ and $w_{o\,H+1}$, equivalent to neural thresholds, where we add formal constant inputs $z_{I+1} = 1$ and $z_{H+1} = 1$ in (1.35) and (1.36), respectively. Respective signal vectors and synaptic weights are connected with one an-

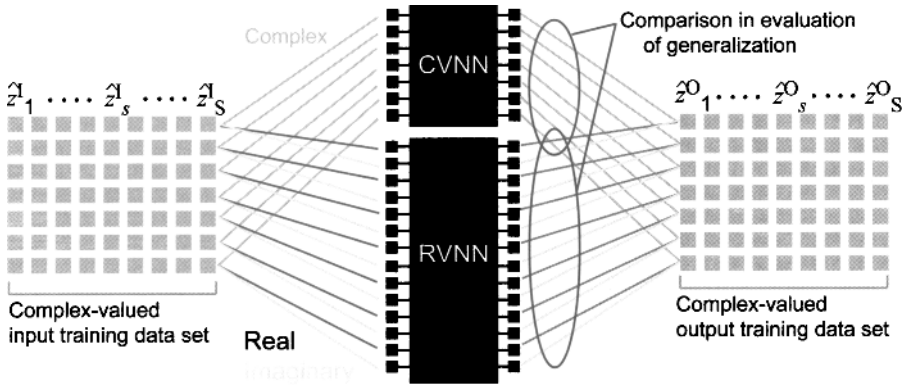


Figure 1.7 Schematic diagram of the learning process for pairs of input–output teachers [35]. (See color insert.)

other through an activation function $f(z)$ as

$$z^{\mathrm{H}} = f\left(\mathbf{W}^{\mathrm{H}} z^{\mathrm{I}}\right) \tag{1.40}$$

$$z^{\mathrm{O}} = f\left(\mathbf{W}^{\mathrm{O}} z^{\mathrm{H}}\right) \tag{1.41}$$

where $f(z)$ is a function of each vector element z ($\in \boldsymbol{C}$) defined as

$$f(z) = \tanh\left(|z|\right) \exp\left(i \arg z\right) \tag{1.42}$$

Figure 1.7 is a diagram to explain the supervised learning process. We prepare a set of teacher signals at the input $\hat{z}_s^{\mathrm{I}} = [\hat{z}_{1s}, ..., \hat{z}_{is}, ..., \hat{z}_{Is}, \hat{z}_{I+1\ s}]^T$ and the output $\hat{z}_s^{\mathrm{O}} = [\hat{z}_{1s}, ..., \hat{z}_{os}, ..., \hat{z}_{Os}]^T$ $(s = 1, ..., s, ...S)$ for which we employ the teacher-signal backpropagation learning. We define an error function E to obtain the dynamics by referring to Refs. [31, 23, 27] as

$$E \equiv \frac{1}{2} \sum_{s=1}^{S} \sum_{o=1}^{O} \left| z_o(\hat{z}_s^{\mathrm{I}}) - \hat{z}_{os} \right|^2 \tag{1.43}$$

$$\left|w_{oh}^{\mathrm{new}}\right| = \left|w_{oh}^{\mathrm{old}}\right| - K \frac{\partial E}{\partial \left|w_{oh}\right|} \tag{1.44}$$

$$\arg w_{oh}^{\mathrm{new}} = \arg w_{oh}^{\mathrm{old}} - K \frac{1}{|w_{oh}|} \frac{\partial E}{\partial (\arg w_{oh})} \tag{1.45}$$

$$\frac{\partial E}{\partial |w_{oh}|} =$$
$$\left(1 - |z_o|^2\right)\left(|z_o| - |\hat{z}_o| \cos\left(\arg z_o - \arg \hat{z}_o\right)\right)|z_h|$$
$$\cdot \cos\left(\arg z_o - \arg \hat{z}_o - \arg w_{oh}\right)$$
$$- |z_o|\,|\hat{z}_o| \sin\left(\arg z_o - \arg \hat{z}_o\right) \frac{|z_h|}{\tanh^{-1}|z_o|}$$
$$\cdot \sin\left(\arg z_o - \arg \hat{z}_o - \arg w_{oh}\right) \tag{1.46}$$

$$\frac{1}{|w_{oh}|} \frac{\partial E}{\partial (\arg w_{oh})} =$$
$$\left(1 - |z_o|^2\right)\left(|z_o| - |\hat{z}_o| \cos\left(\arg z_o - \arg \hat{z}_o\right)\right)|z_h|$$
$$\cdot \sin\left(\arg z_o - \arg \hat{z}_o - \arg w_{oh}\right)$$
$$+ |z_o|\,|\hat{z}_o| \sin\left(\arg z_o - \arg \hat{z}_o\right) \frac{|z_h|}{\tanh^{-1}|z_o|}$$
$$\cdot \cos\left(\arg z_o - \arg \hat{z}_o - \arg w_{oh}\right) \tag{1.47}$$

where $(\cdot)^{\text{new}}$ and $(\cdot)^{\text{old}}$ indicates the update of the weights from $(\cdot)^{\text{old}}$ to $(\cdot)^{\text{new}}$, and K is a learning constant. The teacher signals at the hidden layer $\hat{z}^{\text{H}} = [\hat{z}_1, ..., \hat{z}_h, ..., \hat{z}_H]^T$ is obtained by making the output teacher vector itself $\hat{z}^{\text{O}}$ propagate backward as

$$\hat{z}^{\text{H}} = \left(f\left(\left(\hat{z}^{\text{O}}\right)^* \hat{\mathbf{W}}^{\text{O}}\right)\right)^* \tag{1.48}$$

where $(\cdot)^*$ denotes Hermite conjugate. Using $\hat{z}^{\text{H}}$, the hidden layer neurons change their weights by following (1.44)–(1.47) with replacement of the suffixes o,h with h,i [25, 27].

1.6.1.2 Complex-Valued Neural Network Having Real–Imaginary Separate-Type Activation Function (RI-CVNN) We also investigate the characteristics of complex-valued neural networks having real-imaginary separate-type activation function. Instead of (1.42), a neuron has an activation function expressed as

$$f(z) = \tanh(\mathbf{Re}[z]) + i \tanh(\mathbf{Im}[z]) \tag{1.49}$$

The structure and the dynamics of feedforward processing and backpropagation learning are those described in, for example, Ref. 52.

1.6.1.3 Real-Valued Neural Network Having Double Input Terminals and Output Neurons for Bivariate Procesing (RVNN) Similarly, the forward processing and learning of a RVNN having double input terminals and output neurons are explained as follows. Figure 1.7 includes also this case. We represent a complex number as a pair of real numbers as $z_i = x_{2i-1} + ix_{2i}$. Then we have a double number of terminals for real and imaginary parts of input signals $\boldsymbol{z}_{\text{R}}^{\text{I}}$ and a double number of

output neurons to generate real and imaginary parts of output signals $z_{\mathrm{R}}^{\mathrm{O}}$. We also prepare a double number of hidden neurons for hidden-layer signals $z_{\mathrm{R}}^{\mathrm{H}}$ so that the equivalent number of neurons is the same as that of the above CVNN.

Forward signal processing connects the signal vectors as well as hidden neuron weights $\mathbf{W}_{\mathrm{R}}^{\mathrm{H}}$ and output neuron weights $\mathbf{W}_{\mathrm{R}}^{\mathrm{O}}$ through a real-valued activation function f_{R} as

$$z_{\mathrm{R}}^{\mathrm{I}} = [\overbrace{x_1, \quad x_2}^{\text{real \& imaginary}}, \ldots, x_{2i-1}, x_{2i}, \ldots, x_{2I-1}, x_{2I}, x_{2I+1}, x_{2I+2}]^T \left(= z^{\mathrm{I}}\right) \quad \text{(Input signal vector)} \tag{1.50}$$

$$z_{\mathrm{R}}^{\mathrm{H}} = [x_1, x_2, \ldots, x_{2h-1}, x_{2h}, \ldots, x_{2H-1}, x_{2H}, x_{2H+1}, x_{2H+2}]^T \quad \text{(Hidden-layer output signal vector)} \tag{1.51}$$

$$z_{\mathrm{R}}^{\mathrm{O}} = [x_1, x_2, \ldots, x_{2o-1}, x_{2o}, \ldots, x_{2O-1}, x_{2O}]^T \quad \text{(Output-layer signal vector)} \tag{1.52}$$

$$\mathbf{W}_{\mathrm{R}}^{\mathrm{H}} = [w_{\mathrm{R}hi}] \quad \text{(Hidden neuron weight matrix)} \tag{1.53}$$

$$\mathbf{W}_{\mathrm{R}}^{\mathrm{O}} = [w_{\mathrm{R}oh}] \quad \text{(Output neuron weight mateix)} \tag{1.54}$$

$$z_{\mathrm{R}}^{\mathrm{H}} = f_{\mathrm{R}}\left(\mathbf{W}_{\mathrm{R}}^{\mathrm{H}} z_{\mathrm{R}}^{\mathrm{I}}\right) \tag{1.55}$$

$$z_{\mathrm{R}}^{\mathrm{O}} = f_{\mathrm{R}}\left(\mathbf{W}_{\mathrm{R}}^{\mathrm{O}} z_{\mathrm{R}}^{\mathrm{H}}\right) \tag{1.56}$$

$$f_{\mathrm{R}}(x) = \tanh(x) \tag{1.57}$$

where the thresholds are $w_{\mathrm{R}\,h\,2I+1}$, $w_{\mathrm{R}\,h\,2I+2}$, $w_{\mathrm{R}\,h\,2H+1}$, and $w_{\mathrm{R}\,h\,2H+2}$ with formal additional inputs $x_{2H+1} = 1$, $x_{2H+2} = 1$, $x_{2H+1} = 1$, and $x_{2H+2} = 1$. We employ the conventional error backpropagation learning. That is, we define an error function E_{R} for a set of input and output teacher signals $(\hat{z}_s^{\mathrm{I}}, \hat{z}_s^{\mathrm{O}})$ to obtain the

learning dynamics as

$$E_{\mathrm{R}} \equiv \frac{1}{2}\sum_{s=1}^{S}\sum_{o=1}^{2O}\left|x_o(\hat{z}^{\mathrm{I}}_{\mathrm{R}s}) - \hat{x}_{os}\right|^2 \quad (= E) \tag{1.58}$$

$$w^{\mathrm{new}}_{\mathrm{R}oh} = w^{\mathrm{old}}_{\mathrm{R}oh} - K\frac{\partial E_{\mathrm{R}}}{\partial w_{\mathrm{R}oh}} \tag{1.59}$$

$$w^{\mathrm{new}}_{\mathrm{R}ji} = w^{\mathrm{old}}_{\mathrm{R}hi} - K\frac{\partial E_{\mathrm{R}}}{\partial w_{\mathrm{R}hi}} \tag{1.60}$$

$$\frac{\partial E_{\mathrm{R}}}{\partial w_{\mathrm{R}oh}} = (x_o - \hat{x}_o)\left(1 - x_o^2\right)x_h \tag{1.61}$$

$$\frac{\partial E_{\mathrm{R}}}{\partial w_{\mathrm{R}hi}} = \left(\sum_{o}^{2O}(x_o - \hat{x}_o)\left(1 - x_o^2\right)w_{oh}\right)\left(1 - x_h^2\right)x_i \tag{1.62}$$

1.6.1.4 Dual Real-Valued Neural Networks for Real–Imaginary Separate Processing (dual-RVNN) We consider another type of real-valued neural network in which the real and imaginary parts of input signals are processed separately. It is an extension of dual univariate real-valued neural network having single-layer structure. We may have a variety of ways of mixing and separation of real and imaginary variables in multiple-layer networks. With this network, we examine a completely separate case where the neurons in the real-part network have no connections to those in the imaginary-part network. The learning and processing dynamics are identical to that of the above RVNN except that the numbers of input terminals and output neurons are the same as the CVNNs for the respective real and imaginary networks.

1.6.2 Experimental Setup

1.6.2.1 Simulation Setup Figure 1.8 shows schematically how to observe the generalization characteristics of the networks. We conducted the learning process as follows. We chose the identity mapping as the task to be learned to show the network characteristics most clearly. That is, we take a set of input and output teacher signals as $\hat{z}^{\mathrm{I}}_s = \hat{z}^{\mathrm{O}}_s$ $(s = 1, 2, ..., S)$ with the following conditions. For a signal set showing high coherence, we choose its wavelength in such a manner that a unit wave spans just over the neural input terminals $i = 1, ..., I$, and discrete I points are fed to the network evenly with a constant interval in the unit wave. In more detail, we choose multiple amplitude values between 0 to 1 evenly for $s_{\mathrm{A}} = 1, ..., S_{\mathrm{A}}$ teacher signals as well as multiple time shift amount between 0 to half-wave duration (phase shift in a sinusoidal case between 0 to π) evenly for $s_{\mathrm{t}} = 1, ..., S_{\mathrm{t}}$ teachers. Consequently we generate $S = S_{\mathrm{A}} \times S_{\mathrm{t}}$ points of discrete teacher-signal sets $\hat{z}_{is}$ $(s = 1, 2, ..., S)$ as

$$z_{is} \equiv \frac{s_{\mathrm{A}}}{S_{\mathrm{A}} + 1}\exp\left(i\left(\frac{s_{\mathrm{t}}}{2S_{\mathrm{t}}} + \frac{i}{I}\right)2\pi\right) \tag{1.63}$$

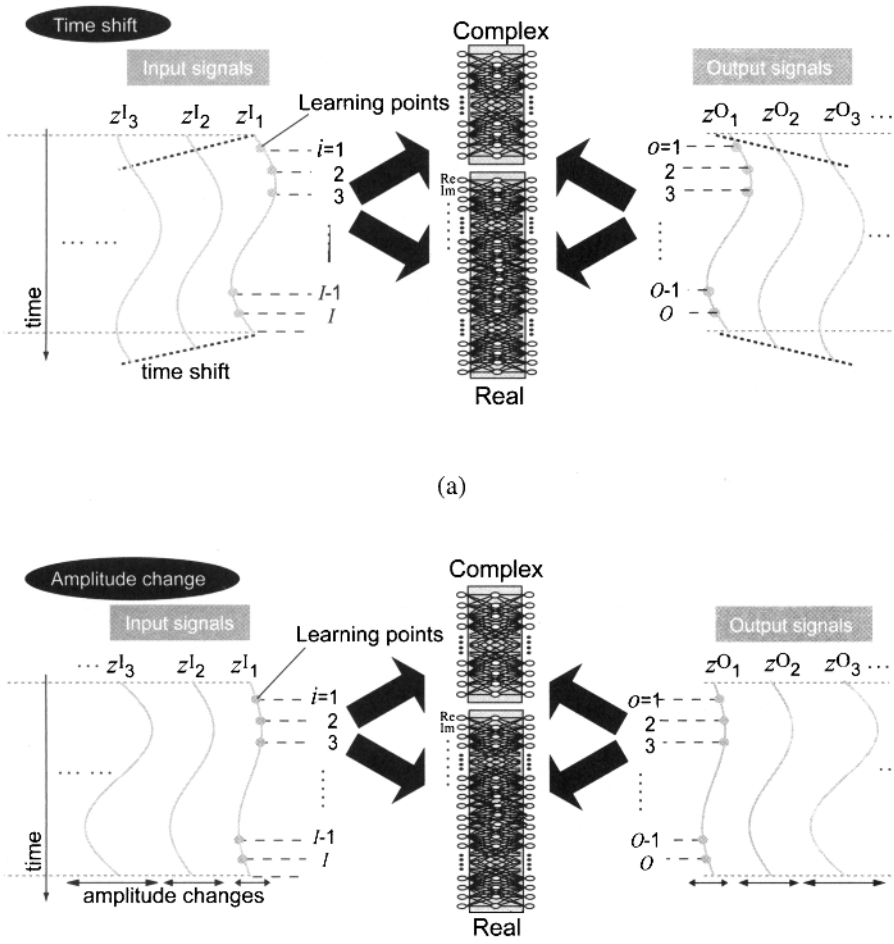
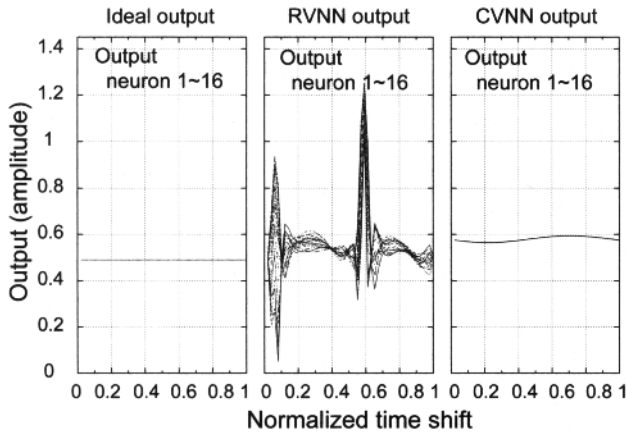


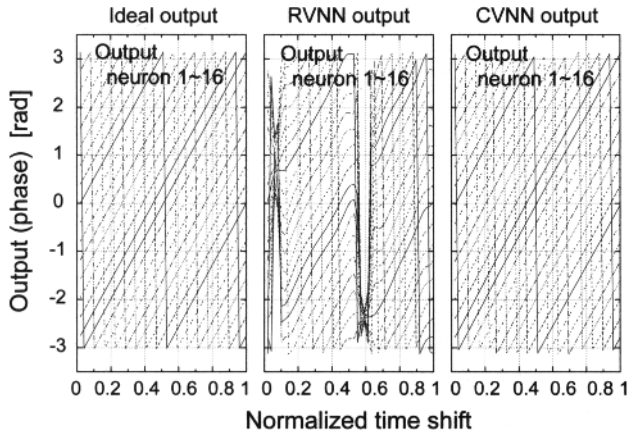
Figure 1.8 Schematic diagrams showing how to feed signals to observe (a) time-shift and (b) amplitude-change generalizations [35]. (See color insert.)

Note that the wavelength, and then the signal frequency, are unchanged. Figure 1.5 includes the manner of the amplitude variation. We add WGN to the sinusoidal wave with various weighting. The noise power is adjusted depending on the signal power and the expected signal-to-noise ratio SNR which is determined in each learning trial.

The dots on the continuous signals in Fig. 1.8 indicates the discrete teacher signal points $\hat{z}_{is}$. We observe the generalization characteristics by inputting signals other than the teachers and evaluate the output errors. Figure 1.8(a) illustrates the observation of outputs when the input signal is shifted in time. The continuous time signal



(a)



(b)

Figure 1.9 Example of (a) amplitude and (b) phase when the input signal gradually shifts in time in the real-valued and complex-valued neural networks (RVNN and CVNN) when no noise is added to sinusoidal signals ($SNR = \infty$) [35].

was generated by the Lagrange interpolation. Figure 1.8(b) shows the observation when the amplitude is changed. We combine the time shift and the amplitude change to evaluate the generalization. In the experiment below, $S_A = 4$, $S_t = 4$, and the neural network parameters are listed in Table 1.1. The learning iteration is 3,000.

1.6.2.2 Heterodyne Signal Experiment We process a heterodyne signal observed in a sonar imaging system. The signal has a carrier of 100 kHz with thermal noise. It is converted into 100 Hz in-phase and quadrature-phase (IQ) intermediate-frequency (IF) signals through an IQ mixer. The imbalance of the IQ mixer is less than 0.3 dB

Table 1.1 Parameters in the neural networks [35]

	CVNN or RI-CVNN	RVNN or dual-RVNN
Number of input neurons	$I = 16$	$2I = 32$
Number of hidden neurons	$H = 25$	$2H = 50$
Number of output neurons	$O = 16$	$2O = 32$
Learning constant	$K = 0.01$	$K = 0.01$

in amplitude and 3 degrees in phase, which is common in this type of systems. The IF signal is recorded by a personal computer (PC) through an analogue/digital converter with 600 k Sample/s sampling frequency. We aim at appropriate interpolation of the signals in time and/or space domain for post-processing to generate high-quality time–space images. When the 100 kHz carrier signal power changes, the SNR also changes for a constant noise power.

1.6.3 Results

1.6.3.1 Examples of Output Signals for Inputs Having Various Coherence Degrees Figure 1.9 displays typical examples of the output signals of the CVNN and RVNN for a single learning trial when $SNR = \infty$, i.e., the signal is completely sinusoidal and coherent. After a learning process, we use other input signal points to investigate the generalization. As mentioned above, the wavelength is adjusted to span over the 16 neural input terminals. For example, we gradually move the input signal forward in time while keeping the amplitude unchanged at $a = 0.5$. Figures 1.9(a) and (b) present the output amplitude and phase, respectively, showing from left-hand side to the right-hand side the ideal output of the identity mapping, the RVNN outputs, and CVNN outputs of the 16 output neurons. The horizontal axes present the time shift t normalized by the unit-wave duration.

In Fig. 1.9(b), we find that the output signals of the RVNN locally deviate greatly from the ideal ones. The learning points are plotted at $t = 0$ (no time shift), where the output amplitude is almost 0.5 for all the neurons. However, with the time course, the amplitude values fluctuate largely. Contrarily, the CVNN amplitude stays almost constant. At the learning point $t = 0$, the value is slightly larger than 0.5, corresponding to the slight nonzero value of the residual error in the learning curve.

In Fig. 1.9(c), the ideal output phase values on the left-hand side exhibit linear increase in time. In the RVNN case, though the phase values at $t = 0$ are the same as those of ideal outputs, the values sometimes swing strongly. In contrast, the CVNN output phase values increase orderly, which is almost identical with the ideal values. In summary, the CVNN presents much better generalization characteristics than the RVNN when the coherence is high, i.e., $SNR = \infty$.

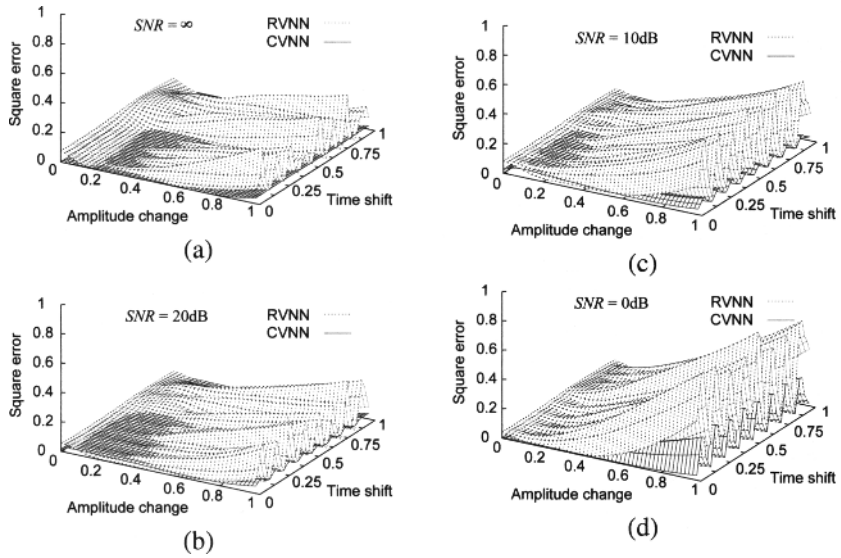


Figure 1.10 Squared generalization errors averaged for 100 trials as functions of amplitude change and time shift for SNR= (a) ∞, (b) 20 dB, (c) 10 dB, and (d) 0 dB, respectively [35].

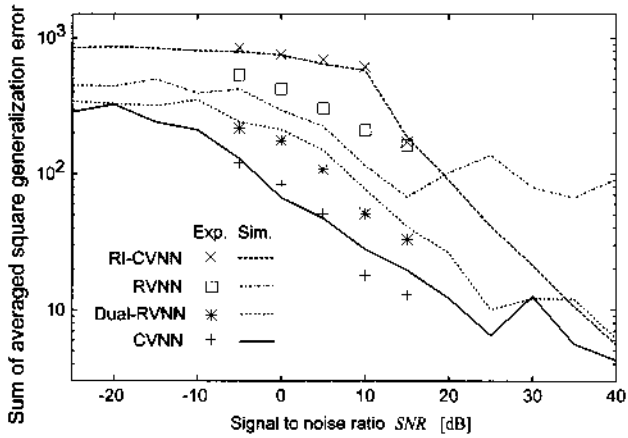


Figure 1.11 Squared generalization errors summed up for all the sampling amplitude-time points shown in Fig. 1.10 versus signal SNR for the real-valued and complex-valued neural networks (CVNN, RI-CVNN, RVNN, and dual-RVNN; curves denote simulations, marks denote experiments) [35].

1.6.3.2 Generalization Error and Its Dependence on the Coherence Here we present statistic results obtained by repeating the above simulations as well as the real-world experiment dealing with the heterodyne signals explained in Section 1.6.2.2.

Figure 1.10 is a three-dimensional representation of the square errors as the average of 100 learning trials for various coherence degree, namely, *SNR*= (a) ∞, (b) 20 dB, (c) 10 dB, and (d) 0 dB, as functions of time shift and amplitude change. The learning points exist at $t = 0$ and amplitude values of $a = 0.2$, 0.4, 0.6, and 0.8. At these points we can find the errors are very small, which corresponds to the almost zero residual errors in the learning curves. However, the errors at the teacher points for lower SNRs are obviously positive. This is because the learning error in some trials fails to converge at zero. As a whole, we notice in Fig. 1.10 that the generalization error of the RVNN are larger than those of the CVNN, in particular in the cases of higher SNR. When SNR is low ($\sim$0 dB), the error of the CVNN also increases.

Figure 1.11 compares quantitatively the generalization errors, summed up for all the sampling amplitude-phase points shown in Fig. 1.10, for the CVNN, RI-CVNN, RVNN, and dual-RVNN as functions of the coherence degree, i.e., SNR. The four curves show the results of the simulation, while the marks indicate experimental results. In all the neural network cases, the generalization error reduces according to the increase of the coherence (increase of SNR). The CVNN curve shows lower errors than other network ones over a wide range of SNR. The dual-RVNN also shows low errors though, at the middle SNR ($SNR = -5$ to 15 dB) the value is 3 to 6 dB larger than that of the CVNN. The error of the simulated RVNN is about 2 dB larger than the dual-RVNN in the low and middle SNR range. The experimental results (marks) of the RVNN are slightly larger. It is remarkable that, in the higher coherence region ($SNR>$ 10 dB), the RVNN curve holds a floor at a nonnegligible level. The RI-CVNN shows a large generalization error in the low coherence region. This is not only because of the errors at non-teacher points but also because of the errors at teacher points. That is, the learning sometimes fails. In the high coherence region ($SNR>$ 20 dB), however, the generalization error decreases and approaches to the curves of the CVNN and dual-RVNN. In summary, we found that the four neural networks present generalization characteristics different among them. The experimental results have been found mostly near to the simulation results. In total, the CVNN shows good generalization characteristics.

1.7 CONCLUSIONS

This chapter first presented recent advances in applications of complex-valued neural networks in various engineering fields, in particular in coherent systems. We also mentioned the history briefly by referring to Parametron. Then we discussed their merits intuitively concerning the degree of freedom in the learning in feedforward layered neural networks as well as the metric specific to the complex-valued networks such as complex inner product. We also considered widely linear systems and circularity not only in data but also in neural dynamics. In the latter part, we

examined the generalization characteristics of complex-valued networks in comparison with real-valued ones. We observed that the complex-valued neural networks show smaller generalization error in the feedforward network to deal with coherent signals. This fact leads to great merits in electronics and engineering fields that deal with wave phenomena and wave-related information such as communications, imaging and sensing, social systems such as traffic signals, frequency-domain processing including frequency-domain multiplexing, and quantum computation and devices. Hypercomplex-valued networks are also promising in the fields related to three-dimensional motion, color processing, and other high-dimensional space information.

REFERENCES

1. I. Aizenberg, E. Myasnikova, M. Samsonova, and J. Reinitz. Temporal classification of drosophila segmentation gene expression patterns by the multi-valued neural recognition method. *Journal of Mathematical Biosciences*, 176(1):145–159, 2002.
2. Igor Aizenberg. *Complex-Valued Neural Networks with Multi-Valued Neurons*. Springer, 2011.
3. Igor Aizenberg, Dimitriy V. Paliy, Jacek M. Zurada, and Jaakko T. Astola. Blur identification by multilayer neural network based on multivalued neurons. *IEEE Transactions on Neural Networks*, 19(5):883–898, May 2008.
4. Takashi Aoyagi, Damri Radenamad, Yukimasa Nakano, and Akira Hirose. Complex-valued self-organizing map clustering using complex inner product in active mmillimeter-wave imaging. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN) 2010 Barcelona*, pages 1346–1351, Barcelona, July 2010. IEEE/INNS.
5. Eduardo Bayro-Corrochano. *Geometric Computing for Wavelet Transforms, Robot Vision, Learning, Control and Action*. Springer, 2010.
6. Ann-Chen Chang, Chih-Wei Jen, and Ing-Jiunn Su. Robust adaptive array beamforming based on independent component analysis with regularized constraints. *IEICE*, E90-B(7):1791–1800, July 2007.
7. Sheng Chen, Lajos Hanzo, and S. Tan. Symmetric complex-valued RBF receiver for muptiple-antenna-aided wireless systems. *IEEE Transactions on Neural Networks*, 19(9):1657–1663, September 2008.
8. Sheng Chen and Xia Hong. Modeling of complex-valued Wiener systems using B-spline neural network. *IEEE Transactions on Neural Networks*, 22(5):818–825, 2011.
9. E. T. Copson. *An Introduction to the Theory of Functions of a Complex Variable*. Oxford: Clarendon Press, 1935.
10. K. L. Du, A. K. Y. Lai, K. K. M. Cheng, and M. N. S. Swamy. Neural methods for antenna array signal processing: A review. *Signal Processing*, 82:547–561, 2002.
11. H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M Koecher, K. Mainzer, J. Neukirch, A. Prestel, and R. Remmert. *Numbers*. Springer-Verlag, (Chapter 3, Section 2), 1983.
12. Q. Gan, P. Saratchandran, N. Sundararajan, and K. R. Subramanian. A complex valued radial basis function network for equalization of fast time varying channels. *IEEE Transactions on Neural Networks*, 10(4):958–960, July 1999.

13. George M. Georgiou and Cris Koutsougeras. Complex domain backpropagation. *IEEE Transactions on Circuits and Systems II*, 39(5):330–334, 1992.

14. S. L. Goh, M. Chen, D. H. Popovic, K. Aihara, and Danilo P. Mandic. Complex valued forecasting of wind profile. *Renewable Energy*, 31:1733–1750, 2006.

15. S. L. Goh and Danilo P. Mandic. Nonlinear adaptive prediction of complex valued non-stationary signals. *IEEE Transactions on Signal Processing*, 53(5):1827–1836, 2005.

16. S.L. Goh and Danilo P. Mandic. An augmented extended kalman filter algorithm for complex-valued recurrent neural networks. *Neural Computation*, 19(4):1–17, 2007.

17. Eiichi Goto. The parametron – A new circuit element which utilizes non-linear reactors (in Japanese). *Paper of Technical Group of Electronic Computers and Nonlinear Theory, IECE*, July 1954.

18. Eiichi Goto. On the application of parametrically excited non-linear resonators (in Japanese). *The Journal of the Institute of Electrical Communication Engineers of Japan (IECE)*, 38(10):2761, October 1955.

19. Abdul Rahman Hafiz, Md. Faijul Amin, and Kazuyuki Murase. Real-time hand gesture recognition using complex-valued neural network (CVNN). In *International Conference on Neural Information Processing (ICONIP) 2011 Shanghai*, pages 541–549, November 2011.

20. Astri Handayani, Andriyan Bayu Suksmono, Tati L. R. Mengko, and Akira Hirose. Blood vessel segmentation in complex-valued magnetic resonance images with snake active contour model. *International Journal of E-Health and Medical Communications*, 1:41–52, 2010.

21. Takahiro Hara and Akira Hirose. Plastic mine detecting radar system using complex-valued self-organizing map that deals with multiple-frequency interferometric images. *Neural Networks*, 17(8-9):1201–1210, November 2004.

22. Akira Hirose. Complex-valued neural networks (Tutorial, International Joint Conference on Neural Networks (IJCNN) 2009 Atlanta) (IEEE Computational Intelligence Society (CIS) Video Archive.

23. Akira Hirose. Continuous complex-valued back-propagation learning. *Electronics Letters*, 28(20):1854–1855, 1992.

24. Akira Hirose. Dynamics of fully complex-valued neural networks. *Electronics Letters*, 28(16):1492–1494, 1992.

25. Akira Hirose. Applications of complex-valued neural networks to coherent optical computing using phase-sensitive detection scheme. *Information Sciences –Applications–*, 2:103–117, 1994.

26. Akira Hirose, editor. *Complex-Valued Neural Networks: Theories and Applications*. World Scientific Publishing Co. Pte. Ltd., 2003.

27. Akira Hirose. *Complex-Valued Neural Networks*. Springer-Verlag, Heidelberg, Berline, New York, 2006.

28. Akira Hirose. Nature of complex number and complex-valued neural networks. *Frontiers of Electrical and Electronic Engineering in China*, 6(1):171–180, 2011.

29. Akira Hirose. Advances in applications of complex-valued neural networks. *Journal of the Society of Instrument and Control Engineering*, 51(4):351–357, April 2012.

30. Akira Hirose, Yasufumi Asano, and Toshihiko Hamano. Developmental learning with behavioral mode tuning by carrier-frequency modulation in coherent neural networks. *IEEE Transactions on Neural Networks*, 17(6):1532–1543, November 2006.

31. Akira Hirose and Rolf Eckmiller. Behavior control of coherent-type neural networks by carrier-frequency modulation. *IEEE Transactions on Neural Networks*, 7(4):1032–1034, 1996.

32. Akira Hirose, Tomoaki Higo, and Ken Tanizawa. Efficient generation of holographic movies with frame interpolation using a coherent neural network. *IEICE Electronics Express*, 3(19):417–423, 2006.

33. Akira Hirose and Tomoyuki Nagashima. Predictive self-organizing map for vector quantization of migratory signals and its application to mobile communications. *IEEE Transactions on Neural Networks*, 14(6):1532–1540, 2003.

34. Akira Hirose and Yukimasa Nakano. Adaptive identification of landmine class by evaluating the total degree of conformity of ring-SOM. *Australian Journal of Intelligent Information Processing Systems*, 12,(1):23–28, 2010.

35. Akira Hirose and Yoshida Shotaro. Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. *IEEE Transactions on Neural Networks and Learning Systems*, 23(4):541–551, April 2012.

36. Deng Jianping, Narasimhan Sundararajan, and P. Saratchandra. Communication channel equalization using complex-valued minimal radial basis fuction neural networks. *IEEE Transactions on Neural Networks*, 13(3):687–696, May 2002.

37. Mitsuru Kawamoto and Yujiro Inouye. Blind deconvolution of MIMO-FIR systems with colored inputs using second-order statistics. *IEICE Transactions on Fundamentals*, E86-A(3):597–604, March 2003.

38. Sotaro Kawata and Akira Hirose. Frequency-multiplexing ability of complex-valued Hebbian learning in logic gates. *International Journal of Neural Systems*, 12(1):43–51, 2008.

39. Mitsunaga Kinjo, Shigeo Sato, Yuuki Nakamiya, and Koji Nakajima. Neuromorphic quantum computation with energy dissipation. *Physical Review A*, 72:052328, 2005.

40. Shin'ichi Koike and Seiichi Noda. Pre-compensation of transmitter nonlinearity with memory effects in digital QAM systems. *IEICE Transactions on Fundamentals*, E87-A(10):2744–2754, October 2004.

41. Danilo P. Mandic and Vanessa Su Lee Goh. *Complex Valued Nonlinear Adaptive Filters – Noncircularity, Widely Linear and Neural Models*. Wiley, April 2009.

42. Danilo P. Mandic, Susanne Still, and Scott C. Douglas. Duality between widely linear and dual channel adaptive filtering. In *IEEE Internatinal Conference on Acoustics, Speech, and Signal Processing 2009 Taipei*, pages 1729–1732, 2009.

43. Soichi Masuyama and Akira Hirose. Walled LTSA array for rapid, high spatial resolution, and phase sensitive imaging to visualize plastic landmines. *IEEE Transactions on Geoscience and Remote Sensing*, 45(8):2536–2543, August 2007.

44. Soichi Masuyama, Kenzo Yasuda, and Akira Hirose. Multiple mode selection of walled-ltsa array elements for high resolution imaging to visualize antipersonnel plastic landmines. *IEEE Geoscience and Remote Sensing Letters*, 5(4):745–749, October 2008.

45. Nobuyuki Matsui, Teijiro Isokawa, Hiromi Kusamichi, Ferdinand Peper, and Haruhiko Nishimura. Quaternion neural network with geometrical operators. *Journal of Intelligent and Fuzzy Systems*, 15:149–164, 2004.

46. R. K. Mishra and A. Patnaik. Neurospectral computation for input impedance of rectangular microstrip antenna. *Electronics Letters*, 35:1691–1693, 1999.

47. R. K. Mishra and A. Patnaik. Designing rectangular patch antenna using the neurospectral method. *IEEE Transactions on Antennas and Propagation*, 51:1914–1921, 2003.

48. Y. Nakamiya, M. Kinjo, O. Takahashi, S. Sato, and K. Nakajima. Quantum neural network composed of kane's qubits. *Japanese Journal of Applied Physics*, 45(10A):8030–8034, October 2006.

49. Yukimasa Nakano and Akira Hirose. Improvement of plastic landmine visualization performance by use of ring-csom and frequency-domain local correlation. *IEICE Transactions on Electronics*, E92-C(1):102–108, January 2009.

50. Ikuko Nishikawa and Yasuaki Kuroe. Dynamics of complex-valued neural networks and its relation to a phase oscillator system. In *International Conference on Neural Information Processing (ICONIP) 2004 Calcutta*, pages 122–129, Berlin, November 2004. Springer.

51. Tomohiro Nishino, Ryo Yamaki, and Akira Hirose. Ultrasonic imaging for boundary shape generation by phase unwrapping with singular-point elimination based on complex-valued Markov random field model. *IEICE Transactions on Fundamentals*, E93-A(1):219–226, January 2010.

52. Tohru Nitta. An extension of the back-propagation algorithm to complex numbers. *Neural Networks*, 10:1391–1415, 1997.

53. Tohru Nitta, editor. *Complex-Valued Neural Networks: Utilizing High–Dimensional Parameters*. Information Science Reference, Pensylvania, February 2009.

54. Tohru Nitta and Yasuaki Kuroe. Special session on Clifford algebraic neural networks. In *International Conference on Neural Information Processing (ICONIP) 2011 Shanghai*. Springer, 2011.

55. Dong-Chul Park and Tae-Kyun Jung Jeong. Complex-bilinear recurrent neural networks for equilization of a digital satellite channel. *IEEE Transactions on Neural Networks*, 13(3):711–725, March 2002.

56. Bernard Picinbono and Pascal Chevalier. Widely linear estimation with complex data. *IEEE Transactions on Signal Processing*, 43(8):2030–2033, August 1995.

57. Shigeo Sato, Mitsunaga Kinjo, and Koji Nakajima. An approach for quantum computing using adiabatic evolution algorithm. *Jpnanese Journal of Appl.ied Physics 1*, 42(11):7169–7173, 2003.

58. H. Sawada, R. Mukai, S. Araki, and S. Makino. Polar coordinate based nonlinear function for frequency-domain blind source separation. *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences*, E86A:590–596, 2003.

59. Andriyan Bayu Suksmono and Akira Hirose. Adaptive noise reduction of insar image based on complex-valued mrf model and its application to phase unwrapping problem. *IEEE Trans. on Geoscience and Remote Sensing*, 40(3):699–709 (followed by publisher's errata on Fig.12), 2002.

60. Andriyan Bayu Suksmono and Akira Hirose. Beamforming of ultra-wideband pulses by a complex-valued spatio-temporal multilayer neural network. *International Journal of Neural Systems*, 15(1):1–7, 2005.

61. Hidetosi Takahasi. An exerimental decimal calculator (in Japanese). *Paper of Technical Group of Electronic Computers, IECE*, March 1956.

62. Chor Shen Tay, Ken Tanizawa, and Akira Hirose. Error reduction in holographic movies using a hybrid learning method in coherent neural networks. *Applied Optics*, 47(28):5221–5228, 2008.

63. Xiaoqiu Wang, Hua Lin, Jianming Lu, and Takashi Yahagi. Combining recurrent neural networks with self-organizing map for channel equalization. *IEICE Transactions on Communications*, E85-B(10):2227–2235, October 2002.

64. Yili Xia, Beth Jelfs, Marc M. Van Hulle, Jose C. Principe, and Danilo P. Mandic. An augumented echo state network for nonlinear adaptive filetering of complex noncircular signals. *IEEE Transactions on Neural Networks*, 22(1):74–83, January 2011.

65. Ryo Yamaki and Akira Hirose. Singular unit restoration in interferograms based on complex-valued Markov random field model for phase unwrapping. *IEEE Geoscience and Remote Sensing Letters*, 6(1):18–22, January 2009.

66. Chih-Chung Yang and N.K. Bose. Landmine detection and classification with complex-valued hybrid neural network using scattering parameters dataset. *IEEE Transactions on Neural Networks*, 16(3):743–753, May 2005.

67. W.H. Yang, K.K. Chan, and P.R. Chang. Complex-valued neural-network for direction-of-arrival estimation. *Electronics Letters*, 30:574–575, 1994.

68. Cheolwoo You and Daesik Hong. Nonlinear blind equalization schemes using complex-valued muptilayer feedforward neural networks. *IEEE Transactions on Neural Networks*, 9:1442–1455, 1998.

69. Yanwu Zhang and Yuanliang Ma. CGHA for principal component extraction in the complex domain. *IEEE Transactions on Neural Networks*, 8:1031–1036, 1997.